Mathematics of Robustness, Risk & Failure

Volume II: Reliability Theory & Failure Processes

Survival analysis, hazard functions, and the mathematics of material and system degradation

---

Every derivation step by step

Every symbol defined

Explicit numerical examples

Geometric intuitions

Silvano Salvador

Contents

5.1 Survival, Hazard, and Cumulative Hazard Functions 3

5.2 The Exponential, Weibull, and Log-Normal Failure Models 7

5.3 The Bathtub Curve and Mixtures of Failure Modes 10

5.4 Kaplan–Meier Estimation and Censored Data 14

5.5 Proportional Hazards and the Cox Model 17

6.1 The Wöhler (S-N) Curve and Basquin's Law 20

6.2 Linear and Nonlinear Cumulative Damage: Miner's Rule and Beyond 24

6.3 Paris' Law and Fatigue Crack Growth 28

6.4 Stochastic Degradation: Gamma and Inverse Gaussian Processes 31

6.5 First-Passage-Time Models for Failure 35

7.1 Renewal Processes and the Renewal Equation 38

7.2 Renewal-Reward Processes and Long-Run Costs 42

7.3 Age and Block Replacement Policies 46

7.4 Imperfect Repair Models and Virtual Age 50

7.5 Competing Risks and Multiple Failure Modes 53

8.1 Structure Functions: Series, Parallel, and k-out-of-n Systems 57

8.2 Reliability Block Diagrams and Fault Trees 60

8.3 Importance Measures: Birnbaum, Criticality, and Fussell–Vesely 65

8.4 Common-Cause Failures and Dependence Modelling 68

8.5 Redundancy Allocation and Optimal Design 72

Bayesian Reliability Analysis 80

Accelerated Life Testing and Reliability Extrapolation 82

Cross-Volume Connections and Methodological Synthesis 83

Numerical Tour: Part 5 — From Raw Data to Survival Curve 86

Formula Reference: Part 5 — Survival Analysis 87

Numerical Tour: Part 6 — Fatigue Life Prediction Pipeline 89

Formula Reference: Part 6 — Fatigue and Degradation 90

Numerical Tour: Part 7 — Optimising a Fleet Maintenance Programme 92

Formula Reference: Part 7 — Renewal and Maintenance 93

Numerical Tour: Part 8 — System Reliability for a Redundant Power Supply 95

Formula Reference: Part 8 — System Reliability 96

Probabilistic Bounds and Approximation Theory in Reliability 99

Exercises and Problems — Part 5 101

Exercises and Problems — Part 6 102

Exercises and Problems — Part 7 103

Exercises and Problems — Part 8 104

Bibliographic Notes and Further Reading 106

Computational Notes: Algorithms and Implementation 108

Glossary of Key Terms and Symbols 110

Open Problems and Research Frontiers 114

Mathematical Foundations 116

Proofs of Selected Theorems 118

Statistical Tables 121

Notation Index 123

Multivariate Reliability and Copula Methods 125

Stochastic Processes in Reliability: A Concise Reference 128

Reliability Data Sources and Industry Standards 132

Extended Case Study: Level 1 PRA for a Simplified Nuclear Safety System 135

Summary of Volume II 137

Preface to Volume II

Scope and motivation. Volume II of the *Mathematics of Robustness, Risk & Failure* series addresses the mathematics of failure in time: when does a component fail? How does failure probability evolve as a system ages, accumulates damage, or degrades? How do we estimate these quantities from data, and how do we design systems and maintenance policies that are optimally robust?

The unifying concept is the *hazard function* $h(t)$ — the instantaneous conditional failure rate given survival to time t. Every model in this volume, from the simplest exponential to the most complex Bayesian joint degradation-survival model, is a specification of how $h(t)$ depends on time, covariates, and history. Understanding $h(t)$ deeply — its mathematical structure, its physical interpretation, its statistical estimation, and its implications for design — is the intellectual core of reliability theory.

Structure. The volume is organised into four parts of five sections each, designed to be read sequentially:

- **Part 5: Survival Analysis and Hazard Functions.** The foundational framework: survival function, hazard, cumulative hazard, the three canonical parametric families (Exponential, Weibull, Log-Normal), the bathtub curve, nonparametric estimation under censoring (Kaplan–Meier, Nelson–Aalen), and the Cox proportional hazards regression model. This part provides the statistical toolkit that underlies all subsequent material.

- **Part 6: Fatigue, Fracture, and Degradation Mechanics.** The physics-based counterpart to the statistical framework. Basquin's S-N law and its probabilistic generalisation (P-S-N curves) govern high-cycle fatigue; Palmgren–Miner's rule and its nonlinear extensions model cumulative damage under variable loading; Paris' law governs fatigue crack growth within linear elastic fracture mechanics; Gamma and Inverse Gaussian stochastic processes model continuous degradation paths; and first-passage-time theory (the Wiener process FPT, Fokker–Planck equations) provides the bridge between degradation models and failure time distributions.

- **Part 7: Renewal Theory and Maintenance Models.** The operational mathematics of repeated failures and repairs: the renewal equation and its asymptotic theory (the elementary and key renewal theorems, the central limit theorem for renewals), the renewal-reward theorem as the universal tool for long-run cost analysis, age and block replacement policies and their optimisation, imperfect repair models (Kijima's virtual age, the geometric process, the Crow–AMSAA reliability growth model), and competing risks (cause-specific hazards, cumulative incidence functions, the Aalen–Johansen estimator, Fine–Gray regression).

- **Part 8: System Reliability and Redundancy.** From components to systems: structure functions, minimal path and cut sets, the reliability polynomial, fault trees, reliability block diagrams, component importance measures (Birnbaum, Fussell–Vesely, RAW, RRW), common-cause failure modelling (β-factor, α-factor, copula models), and the redundancy allocation problem (dynamic programming, Lagrangian relaxation, standby redundancy). The part concludes with reliability-centred maintenance (RCM) as the systematic integration of all preceding tools.

Mathematical prerequisites. The reader is assumed to be familiar with probability theory and mathematical statistics at the graduate level (random variables, distributions, maximum likelihood estimation, hypothesis testing), introductory stochastic processes (Poisson

processes, Markov chains), and basic calculus and linear algebra. Volume I's coverage of extreme value theory and tail distributions is referenced where relevant.

Notation conventions. Throughout this volume:

- T denotes a non-negative random variable representing *time to failure*.
- $F(t) = \Pr(T \leq t)$: cumulative distribution function (CDF) of T.
- $S(t) = 1 - F(t) = \Pr(T > t)$: survival function.
- $f(t) = F'(t)$: probability density function (pdf).
- $h(t) = f(t)/S(t)$: hazard (failure rate) function.
- $H(t) = -\ln S(t)$: cumulative hazard function.
- $\delta_i \in \{0, 1\}$: censoring indicator (1 = observed failure, 0 = censored).
- $\hat{\cdot}$: MLE or nonparametric estimate.
- $\mathbf{1}[\cdot]$: indicator function.
- $\phi(\cdot)$, $\Phi(\cdot)$: standard normal pdf and CDF.

Each section follows a consistent structure: (1) a formal Definition Box establishing the mathematical objects, (2) derivations from first principles with no steps omitted, (3) at least one numerical Example Box with explicit computations, and (4) extensions and connections to other parts of the volume and the series.

PART 5

Survival Analysis and Hazard Functions

Core formalism for time-to-failure: survival functions, hazard rates, and the classical parametric failure-time families.

5.1 Survival, Hazard, and Cumulative Hazard Functions

Definition: Survival and Hazard Functions

Let $T \geq 0$ be a continuous random variable representing the *time to failure* of a component or system. Its distribution is characterised by three equivalent representations:

- **CDF**: $F(t) = \Pr(T \leq t)$, the probability of failure by time t.
- **Survival function**: $S(t) = \Pr(T > t) = 1 - F(t)$.
- **Hazard function**: $h(t) = \lim_{\Delta t \to 0} \dfrac{\Pr(t < T \leq t + \Delta t \mid T > t)}{\Delta t} = \dfrac{f(t)}{S(t)}$, where $f(t) = F'(t)$ is the density.
- **Cumulative hazard**: $H(t) = \int_0^t h(u)\, du = -\ln S(t)$.

Derivation of the hazard–survival link. From $H(t) = -\ln S(t)$ it follows directly that

$$S(t) = \exp(-H(t)) = \exp\left(- \int_0^t h(u)\, du \right).$$

Differentiating: $f(t) = -S'(t) = h(t)S(t)$, confirming $h(t) = f(t)/S(t)$.

Interpretation of $h(t)$. The hazard $h(t)\,\Delta t$ is approximately the conditional probability that a component which has survived to t will fail in the next small interval $(t, t+\Delta t]$. It is an instantaneous failure rate, *not* a probability in itself, and can exceed 1 for $\Delta t = 1$ (unlike a probability). Units: failures per unit time.

Mean residual life (MRL). For a component of age t, the expected additional lifetime is

$$e(t) = \mathbb{E}[T - t \mid T > t] = \frac{\int_t^\infty S(u)\,du}{S(t)}.$$

The MRL completely determines the distribution: given $e(t)$, one recovers $S(t) = \frac{e(0)}{e(t)} \exp\!\Big(-\int_0^t \frac{du}{e(u)}\Big)$.

The hazard function as a modelling primitive. In reliability, $h(t)$ is often specified directly (rather than f or S) because its shape encodes the failure mechanism:

- *Constant* $h(t) = \lambda$: memoryless, random failures (exponential distribution).
- *Increasing* $h(t)$: wear-out failures — older components fail faster.
- *Decreasing* $h(t)$: infant mortality — early failures dominate.
- *Bathtub-shaped* $h(t)$: all three phases (discussed in Section 5.3).

Example — Computing S, H, and h for a Weibull distribution

Let $T \sim \text{Weibull}(\lambda, k)$ with CDF $F(t) = 1 - e^{-(\lambda t)^k}$ for $t \geq 0$.
Survival: $S(t) = e^{-(\lambda t)^k}$.
Density: $f(t) = k\lambda^k t^{k-1} e^{-(\lambda t)^k}$.
Hazard: $h(t) = \frac{f(t)}{S(t)} = k\lambda^k t^{k-1}$.
Cumulative hazard: $H(t) = (\lambda t)^k$.
Numerical instance. Take $\lambda = 0.1$, $k = 2$ (wear-out regime). At $t = 5$:

$$\begin{aligned} S(5) &= e^{-(0.5)^2} = e^{-0.25} \approx 0.779, \\ h(5) &= 2(0.1)^2 \cdot 5 = 0.10 \text{ (failures/unit time)}, \\ H(5) &= (0.1 \cdot 5)^2 = 0.25. \end{aligned}$$

At $t = 10$: $h(10) = 0.20$, double the hazard at $t = 5$ — the increasing hazard correctly models accelerating wear-out.

Discrete-time analogue. If T takes values in $\{1, 2, \ldots\}$, the hazard at time j is $h_j = \Pr(T = j \mid T \geq j)$, and $S(t) = \prod_{j \leq t}(1 - h_j)$. This discrete form underlies the Kaplan–Meier estimator (Section 5.4).

Transformed time and TTT plots. The *total-time-on-test* (TTT) transform $\phi(u) = H^{-1}(-\ln(1-u))/\mu$ for $u \in [0,1]$ (where $\mu = \mathbb{E}[T]$) is convex for IFR distributions, concave for DFR, and linear for exponential. Plotting the empirical TTT transform gives a nonparametric diagnostic of the hazard shape.

Hazard function classes and their preservation. A distribution F belongs to the *increasing failure rate* (IFR) class if $h(t)$ is non-decreasing; it belongs to the *increasing failure rate average* (IFRA) class if $H(t)/t = -\ln S(t)/t$ is non-decreasing. The IFRA class is closed under formation of coherent system lifetimes (Barlow–Proschan theorem), a key property exploited in system reliability.

IFRA Closure Theorem (Barlow–Proschan, 1975)

If component lifetimes $T_1, \dots, T_n$ are i.i.d. IFRA, then the lifetime of any coherent system formed from these components is also IFRA.

Proof sketch. For a coherent structure function ϕ, the system survival is $S_s(t) = h(\mathbf{S}(t))$ where h is the *reliability function* (polynomial evaluated at component reliabilities). Since $S_i(t) = e^{-H_i(t)}$ and $H_i(t)/t$ is non-decreasing (IFRA), one shows that $-\ln S_s(t)/t$ is also non-decreasing by convexity arguments on the reliability polynomial.

Laplace transform of the failure time. The moment generating function of T is $\mathcal{M}(s) = \mathbb{E}[e^{-sT}] = \tilde{f}(s) = \int_0^\infty e^{-st} f(t)\, dt$, the Laplace transform of the density. For the Weibull: no closed form except $k = 1$ (exponential, $\mathcal{M}(s) = \lambda/(\lambda + s)$) and $k = 2$ (Rayleigh, involves the complementary error function). The moments are:

$$\mathbb{E}[T^r] = \frac{\Gamma(1 + r/k)}{\lambda^r}, \quad r > 0.$$

Right-truncated and interval-censored data. Beyond right censoring, reliability data often appear as:

- *Left truncation*: a component enters the study only if it has already survived to age l_i; the likelihood factor is $f(t_i)/S(l_i)$.
- *Interval censoring*: only known that $T_i \in (L_i, U_i]$; contribution $S(L_i) - S(U_i)$.
- *Exact observations*: contribution $f(t_i)$.

The full likelihood is $L = \prod_i f(t_i)^{\delta_i} [S(L_i) - S(U_i)]^{\gamma_i} S(t_i)^{1-\delta_i-\gamma_i} / S(l_i)$, maximised jointly.

Proportional odds model. An alternative to the PH model is the proportional odds (PO) model:

$$\frac{F(t \mid \mathbf{x})}{S(t \mid \mathbf{x})} = \frac{F_0(t)}{S_0(t)} \exp(\boldsymbol{\beta}^\top \mathbf{x}),$$

implying that the covariate effect on the odds of failure decreases over time — appropriate when early failures are strongly covariate-dependent but long-term survivors are similar across groups. The PO model is nested in the Burr family of distributions.

Quantile regression for survival data. The τ-th quantile regression model specifies $Q_\tau(t \mid \mathbf{x}) = Q_\tau^0(t) + \boldsymbol{\beta}_\tau^\top \mathbf{x}$, where Q_τ^0 is the baseline quantile function. Unlike PH, quantile regression directly models the effect of covariates on specific percentiles of the failure-time distribution, without a global proportionality assumption.

Example — Interval censoring in a shock accelerometer test

Twenty accelerometers are inspected only at $t = 500$, 1000, 2000 h. At $t = 500$: 2 failed. At $t = 1000$: 5 more. At $t = 2000$: 6 more. Remaining 7: right-censored at 2000 h. Under a Weibull assumption, the interval-censored log-likelihood is:

$$\ell = 2 \ln F(500) + 5 \ln[F(1000) - F(500)] + 6 \ln[F(2000) - F(1000)] + 7 \ln S(2000).$$

MLE via grid search: $\hat{k} = 1.72$, $\hat{\lambda} = 1/1350$ h^{-1}. Predicted median life: $t_{0.5} = (\ln 2)^{1/k}/\hat{\lambda} = (0.693)^{0.581} \times 1350 \approx 1070$ h.

Accelerated Failure Time (AFT) models. The AFT model specifies:

$$\ln T = \boldsymbol{\beta}^\top \mathbf{x} + \sigma \epsilon,$$

where ϵ has a fixed distribution (extreme value gives Weibull AFT; normal gives log-normal AFT; logistic gives log-logistic AFT). The covariate $\mathbf{x}$ *scales* the time axis: $S(t \mid \mathbf{x}) = S_0(t/\psi(\mathbf{x}))$ where $\psi(\mathbf{x}) = \exp(\boldsymbol{\beta}^\top \mathbf{x})$ is the acceleration factor. AFT models are fully parametric, easily interpretable, and give direct quantile predictions.

Connection between AFT and PH. The AFT and PH models coincide only for the Weibull family: if $T \sim \text{Weibull}(\lambda, k)$ and covariates act multiplicatively on λ, then $h(t \mid \mathbf{x}) = h_0(t)e^{k\boldsymbol{\beta}^\top \mathbf{x}}$ (PH) and $\ln T = -\boldsymbol{\beta}^\top \mathbf{x} + \sigma\epsilon$ (AFT) simultaneously hold. For all other distributions, AFT and PH imply different covariate effects.

Hazard rate estimation: kernel smoothing. The Nelson–Aalen estimator gives a step-function estimate of $H(t)$. A smooth hazard estimate is obtained by differentiating a kernel-smoothed version:

$$\hat{h}(t) = \frac{1}{b}\sum_{j=1}^{m} K\left(\frac{t - t_j}{b}\right)\frac{d_j}{n_j},$$

where K is a kernel function (Epanechnikov: $K(u) = 0.75(1-u^2)_+$; boundary-corrected at $t = 0$) and b is the bandwidth. The optimal bandwidth b^* minimises the integrated mean squared error (IMSE):

$$\text{IMSE}(b) = \int \text{Bias}[\hat{h}(t)]^2 + \text{Var}[\hat{h}(t)]\, dt \approx C_1/b^4 + C_2 b/n.$$

Cross-validation bandwidth: $b^* = \arg\min_b \widehat{\text{IMSE}}(b)$ via leave-one-out log-likelihood.

Nonparametric maximum likelihood under shape constraints. If $h(t)$ is constrained to be monotone non-decreasing (IFR), the NPMLE is piecewise constant and computed by isotonic regression applied to the Nelson–Aalen increments. The Brunk estimator gives: $\hat{h}_j = \max_{k \le j} \min_{l \ge j} \frac{\sum_{i=k}^{l} d_i/n_i}{t_l - t_{k-1}}$ (a pool-adjacent-violators type result).

Score test for IFR vs. constant hazard. Under $H_0 : h(t) = \lambda$ (exponential) vs. $H_1 : h(t)$ non-decreasing, the score test statistic is the sum of Schoenfeld-type residuals weighted by time: $T = \sum_j t_j(d_j - n_j\hat{\lambda}\Delta t_j)$, which is asymptotically normal under H_0.

Example — Hazard rate estimation for compressor seal failures

$n = 85$ compressor seals, 23 failures over 5 years (right-censored data). Nelson–Aalen estimate gives a clearly increasing cumulative hazard. Kernel hazard estimate (Epanechnikov, $b = 0.4$ years): $\hat{h}(1) = 0.028$, $\hat{h}(3) = 0.071$, $\hat{h}(5) = 0.118$ failures/year. The hazard roughly triples from year 1 to year 5, consistent with Weibull $k \approx 2$. Parametric Weibull MLE: $\hat{k} = 1.97$, $\hat{\lambda} = 0.053$/year. Kolmogorov–Smirnov test: $D = 0.074$, $p = 0.64$ (good fit). The Weibull is used for extrapolation to $t = 10$ years where no data exist.

Hazard function for interval data via EM. For interval-censored data (L_i, U_i), the NPMLE of F places probability mass only at the *inner intervals* of the Turnbull set. The EM algorithm:

- E-step: allocate expected probability mass $p_{ij} \propto q_j \mathbf{1}[s_j \in (L_i, U_i)]$ to each interval.
- M-step: $\hat{q}_j = n^{-1}\sum_i p_{ij}$ (average allocation).

Convergence to the nonparametric MLE is guaranteed; the convergence rate is $O(1/k)$ (slow), so acceleration via Aitken Δ^2 is common.

Hazard regression via splines. Flexible parametric models for $h_0(t)$ use cubic B-splines: $\ln h_0(t) = \sum_{l=1}^{L} \gamma_l B_l(t)$, where B_l are B-spline basis functions and L controls smoothness.

Penalised partial likelihood:

$$\ell_{\text{pen}}(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \ell(\boldsymbol{\gamma}, \boldsymbol{\beta}) - \frac{\nu}{2} \int [h_0''(t)]^2 \, dt,$$

with smoothing parameter ν chosen by AIC or GCV. This gives a smooth, data-driven hazard shape without the proportionality constraint.

5.2 The Exponential, Weibull, and Log-Normal Failure Models

The Three Core Parametric Families

Exponential ($\lambda > 0$): $h(t) = \lambda$ (constant), $S(t) = e^{-\lambda t}$, $\mathbb{E}[T] = 1/\lambda$. The unique *memoryless* continuous distribution.
Weibull ($\lambda > 0$, $k > 0$): $h(t) = k\lambda(\lambda t)^{k-1}$, $S(t) = e^{-(\lambda t)^k}$. IFR if $k > 1$, DFR if $k < 1$, exponential if $k = 1$.
Log-Normal ($\mu \in \mathbb{R}$, $\sigma > 0$): $T = e^{\mu + \sigma Z}$ with $Z \sim \mathcal{N}(0, 1)$.

$$h(t) = \frac{\phi\left(\frac{\ln t - \mu}{\sigma}\right)}{\sigma t \, \Phi\left(-\frac{\ln t - \mu}{\sigma}\right)},$$

where ϕ and Φ are the standard normal pdf and CDF. The hazard is *unimodal* (rises then falls), appropriate for many electronic and biological failure modes.

Memorylessness of the exponential. For $T \sim \text{Exp}(\lambda)$:

$$\Pr(T > s + t \mid T > s) = \Pr(T > t) \quad \forall \, s, t \geq 0.$$

Proof. $\Pr(T > s + t \mid T > s) = \frac{S(s+t)}{S(s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \Pr(T > t)$. ∎

The memoryless property means a component's remaining lifetime is independent of how long it has already operated. This is realistic only for failure modes dominated by external shocks (e.g. cosmic ray bit-flips in semiconductors), not for wear.

Weibull shape parameter k.

- $k < 1$: DFR — early failures ("infant mortality"), defective units fail first.
- $k = 1$: Exponential — random, shock-driven failures.
- $k = 2$: Rayleigh — linear hazard, common in mechanical wear.
- $k \approx 3.5$: Nearly symmetric, approximates a normal distribution.
- $k > 1$ general: IFR — progressive wear-out.

Fréchet connection. The Weibull distribution is max-stable under minimum: if $T_1, \ldots, T_n \overset{\text{iid}}{\sim} \text{Weibull}(\lambda, k)$, then $\min_i T_i \sim \text{Weibull}(n^{1/k}\lambda, k)$ (after normalisation). This is the counterpart of the GEV theory for minima (Vol. I, Part 3).

Log-Normal and fatigue. The log-normal arises naturally from the *multiplicative degradation* model: if the total damage at failure is $D = \prod_{i=1}^{n} X_i$ where X_i are i.i.d. random increments, then by the CLT applied to $\ln D = \sum \ln X_i$, $\ln D$ is approximately normal. This explains the prevalence of the log-normal in corrosion, crack propagation, and biological growth models.

Generalised Gamma. The three-parameter Generalised Gamma $\text{GG}(\lambda, k, \gamma)$ with density $f(t) = \frac{\gamma \lambda^{\gamma} t^{\gamma k - 1} e^{-(\lambda t)^{\gamma}}}{\Gamma(k)}$ nests exponential ($k = 1, \gamma = 1$), Weibull ($k = 1$), Gamma ($\gamma = 1$), and

log-normal ($k \to \infty$ after reparametrisation). It is the most flexible single-mode parametric family for failure times.

Example — MLE for the Weibull distribution

Given n uncensored failure times $t_1, \dots, t_n$, the log-likelihood is:

$$\ell(\lambda, k) = n \ln k + nk \ln \lambda + (k-1) \sum_{i=1}^{n} \ln t_i - \lambda^k \sum_{i=1}^{n} t_i^k.$$

Setting $\partial \ell / \partial \lambda = 0$: $\hat{\lambda}^k = n / \sum_i t_i^k$, giving $\hat{\lambda} = \left(n / \sum t_i^k\right)^{1/k}$.
Setting $\partial \ell / \partial k = 0$ yields the implicit equation:

$$\frac{1}{\hat{k}} + \frac{\sum_{i=1}^{n} \ln t_i}{n} - \frac{\sum_{i=1}^{n} t_i^{\hat{k}} \ln t_i}{\sum_{i=1}^{n} t_i^{\hat{k}}} = 0,$$

solved numerically (e.g. Newton–Raphson). **Numerical instance.** Data: $t = (2, 3, 5, 7, 11)$ hours. Initial guess $k_0 = 1$. After 5 Newton iterations: $\hat{k} \approx 1.83$, $\hat{\lambda} \approx 0.149$. The shape $\hat{k} > 1$ indicates increasing failure rate — consistent with a wear-out process.

Accelerated life testing (ALT). Under an accelerated stress $s > s_0$ (elevated temperature, voltage, etc.), the Arrhenius model posits $\lambda(s) = \lambda_0 \exp(-E_a / k_B s)$, where E_a is activation energy and k_B is Boltzmann's constant. For the Weibull family, stress enters multiplicatively: $T(s) \overset{d}{=} T(s_0) \cdot (s_0/s)^{\beta}$ (power-law acceleration). This gives a log-linear model for $\ln T$ on $\ln s$, enabling extrapolation from accelerated test data to normal operating conditions.

Gamma and Gompertz distributions. Beyond the three canonical families:

Gamma(α, β): $f(t) = \beta^\alpha t^{\alpha-1} e^{-\beta t} / \Gamma(\alpha)$. Hazard: non-decreasing for $\alpha > 1$ (IFR), non-increasing for $\alpha < 1$ (DFR), constant for $\alpha = 1$ (exponential). No closed-form $S(t)$ (requires incomplete gamma). MLE: $\hat{\beta} = \alpha / \bar{t}$; $\hat{\alpha}$ solves $\ln \hat{\alpha} - \psi(\hat{\alpha}) = \ln \bar{t} - \overline{\ln t}$ where ψ is the digamma function.

Gompertz: $h(t) = \lambda e^{\gamma t}$, giving $S(t) = \exp(-(\lambda/\gamma)(e^{\gamma t} - 1))$. The hazard is exponentially increasing — widely used in actuarial science (human mortality) and in biological ageing models. The Makeham extension adds a constant term: $h(t) = \lambda e^{\gamma t} + \mu$ (Makeham's law), separating ageing-related from accidental mortality.

Log-Logistic Distribution

$T \sim \text{LogLogistic}(\alpha, \beta)$: $F(t) = \frac{(\alpha t)^\beta}{1+(\alpha t)^\beta}$, $S(t) = \frac{1}{1+(\alpha t)^\beta}$, $h(t) = \frac{\alpha\beta(\alpha t)^{\beta-1}}{1+(\alpha t)^\beta}$.
The hazard is unimodal (increases then decreases) for $\beta > 1$ and monotone decreasing for $\beta \leq 1$. Closed-form $S(t)$ makes it attractive for interval-censored data. Its proportional odds (PO) model analog:

$$\frac{F(t \mid \mathbf{x})}{S(t \mid \mathbf{x})} = \frac{F_0(t)}{S_0(t)} e^{\boldsymbol{\beta}^\top \mathbf{x}}.$$

Parameter estimation for the Log-Normal. Given $T_i = e^{Y_i}$ with $Y_i \sim \mathcal{N}(\mu, \sigma^2)$, the MLE is simply applied to the logs: $\hat{\mu} = \overline{\ln T}$, $\hat{\sigma}^2 = \overline{(\ln T - \hat{\mu})^2}$. Under right censoring with indicator

δ_i, the censored log-likelihood is:

$$\ell(\mu,\sigma^2)=\sum_i\left[\delta_i\ln\phi\!\left(\frac{\ln t_i-\mu}{\sigma}\right)+(1-\delta_i)\ln\Phi\!\left(\frac{-(\ln t_i-\mu)}{\sigma}\right)\right]-\delta_i\ln(\sigma t_i),$$

solved by EM (E-step: impute censored $\ln T_i$; M-step: standard normal MLE on augmented data).

Model selection criteria. For nested Weibull vs. Exponential ($k=1$ is a boundary): use the likelihood ratio test $\Lambda=2(\ell_W-\ell_E)\sim\chi^2_1$ (boundary of parameter space requires Chernoff's theorem). For non-nested comparisons (Weibull vs. Log-Normal): use AIC $=-2\ell+2p$, BIC $=-2\ell+p\ln n$, or the Vuong likelihood ratio test. In practice, probability plots (Weibull plot: $\ln(-\ln S)$ vs. $\ln t$; log-normal plot: $\Phi^{-1}(F)$ vs. $\ln t$) provide visual diagnostics.

Quantile–quantile (Q-Q) diagnostic. For a hypothesised Weibull$(\hat\lambda,\hat k)$, plot theoretical quantiles $F^{-1}(p_i)$ against empirical quantiles $\hat t_{(i)}$ (order statistics). Departure from the 45° line indicates model misfit. The Anderson–Darling statistic $A^2=-n-\sum_{i=1}^{n}(2i-1)\ln[F(\hat t_{(i)})(1-F(\hat t_{(n+1-i)})]/n$ is the most powerful omnibus test for right-tail behaviour.

Example — Weibull vs log-normal for ball bearing fatigue

Dataset: 23 ball bearing fatigue lives (millions of revolutions). KM estimate. AIC: Weibull = 142.3, Log-Normal = 140.8, Gamma = 141.5. Log-Normal gives marginally best fit. However, Weibull probability plot shows good linearity except for the two lowest order statistics, suggesting a small sub-population of defective bearings (mixture model). Engineering judgement favours Weibull for the main population with a 2-component mixture (defective fraction $\hat p_1=0.06$).

Mixture cure models. In some failure populations a fraction π of units are *immune* (will never fail within the study horizon). The mixture cure model:

$$S(t)=\pi+(1-\pi)S_{\text{susceptible}}(t),$$

where $S_{\text{susceptible}}(t)\to 0$ as $t\to\infty$. The KM plateau (if $\hat S(t^*)>0$ at the last failure time) is evidence for a cure fraction. MLE via EM: E-step imputes the probability that each long-term survivor is cured; M-step updates $\hat\pi$ and the susceptible distribution.

Method of moments estimation. For the Weibull distribution:

$$\mathbb{E}[T]=\lambda^{-1}\Gamma(1+1/k),$$
$$\mathbb{E}[T^2]=\lambda^{-2}\Gamma(1+2/k).$$

Setting sample moments $\hat\mu_1=\bar T$, $\hat\mu_2=\overline{T^2}$ equal to theoretical moments gives a system of two equations in (k,λ). The coefficient of variation $\text{CV}=\sqrt{\text{Var}[T]}/\mathbb{E}[T]=\sqrt{\Gamma(1+2/k)/\Gamma(1+1/k)^2-1}$ is a monotone function of k alone, giving a single-equation estimator: $\hat k=\hat k(\widehat{\text{CV}})$ (tabulated or numerically inverted). Then $\hat\lambda=\Gamma(1+1/\hat k)/\bar T$.

Probability weighted moments (PWM). PWMs $\beta_r=\mathbb{E}[T\{F(T)\}^r]$ are estimated without censoring by $\hat\beta_r=n^{-1}\sum_{i=1}^{n}T_{(i)}\binom{i-1}{r}/\binom{n-1}{r}$ (Landwehr–Wallis–Matalas). For the GEV distribution (relevant for extremes in Vol. I), PWM estimators are available in closed form and more robust than MLE for small samples.

Competing criteria for failure time distributions.

Family	Parameters	IFR	DFR	UBT	Closed $S(t)$
Exponential	1	–	–	–	Yes
Weibull	2	$k > 1$	$k < 1$	–	Yes
Log-Normal	2	–	–	Yes	No
Gamma	2	$\alpha > 1$	$\alpha < 1$	–	No
Log-Logistic	2	$\beta \leq 1$	$\beta \leq 1$	$\beta > 1$	Yes
Gompertz	2	always	–	–	Yes
Burr XII	2	varies	varies	Yes	Yes

Total time on test (TTT) transform. The scaled TTT statistic:

$$\phi_n(k/n) = \frac{\sum_{i=1}^{k} T_{(i)} + (n-k)T_{(k)}}{\sum_{i=1}^{n} T_{(i)}},$$

plotted against k/n, is:

- Concave (above diagonal): IFR.
- Convex (below diagonal): DFR.
- Bathtub-shaped (convex then concave): BT hazard.
- Diagonal: exponential.

The TTT plot is a powerful nonparametric graphical tool for identifying the hazard shape before fitting a parametric model.

Example — TTT plot for hydraulic hose failures

$n = 30$ failure times (hours), sorted. Empirical TTT at $k/n = 0.5$ (15th order statistic): $\phi_n(0.5) = 0.31 < 0.5$ (below diagonal). At $k/n = 0.8$: $\phi_n(0.8) = 0.62 < 0.8$ (still below diagonal). The TTT plot is convex throughout — consistent with DFR (Weibull $k < 1$). Fitted: $\hat{k} = 0.73$, $\hat{\lambda} = 1/380$ h. Interpretation: early-failure-dominated hose failures (material defects prevalent in the batch).

Power-law transformations. If $T \sim \text{Weibull}(\lambda, k)$, then $T^k \sim \text{Exp}(\lambda^k)$. This transformation converts any Weibull sample to an exponential sample, enabling exact tests. The exponential probability plot (plotting $-\ln(1-\hat{F}_i)$ vs. $T_{(i)}^{\hat{k}}$) should be linear; deviations indicate model misspecification.

5.3 The Bathtub Curve and Mixtures of Failure Modes

The Bathtub Hazard

The canonical empirical observation in hardware reliability is that the hazard function $h(t)$ follows a *bathtub curve* with three distinct phases:

1. **Burn-in (infant mortality)**: $h(t)$ decreasing. Caused by manufacturing defects, poor assembly, or material flaws that fail quickly under operating stress. Duration: hours to weeks.
2. **Useful life**: $h(t) \approx$ constant. Random, independent failures dominate; exponential distribution is a good model.
3. **Wear-out**: $h(t)$ increasing. Material degradation, fatigue, corrosion, and creep accumulate; Weibull with $k > 1$ is appropriate.

Mixture model for the bathtub. A standard approach is to model the failure-time distribution as a mixture of two (or three) populations:

$$f(t) = p_1 f_1(t) + p_2 f_2(t) + p_3 f_3(t), \quad p_1 + p_2 + p_3 = 1,$$

where f_1 is a DFR component (e.g. Weibull $k_1 < 1$ for infant mortality), f_2 exponential (useful life), and f_3 IFR (e.g. Weibull $k_3 > 1$ for wear-out). The resulting mixture hazard is

$$h(t) = \frac{p_1 f_1(t) + p_2 f_2(t) + p_3 f_3(t)}{p_1 S_1(t) + p_2 S_2(t) + p_3 S_3(t)}.$$

Even though each f_i has a monotone hazard, the mixture can produce a bathtub shape.

Exponentiated Weibull. Mudholkar and Srivastava (1993) proposed $F(t) = \left[1 - e^{-(\lambda t)^k}\right]^\alpha$. This model has bathtub hazard when $\alpha k < 1$ and unimodal hazard for other parameter ranges, providing a parametric bathtub model without mixing.

IDB (Increasing–Decreasing–Bathtub) taxonomy. Let $h'(t)$ denote the derivative of the hazard. Define:

- IFR (Increasing Failure Rate): $h'(t) > 0$ for all t.
- DFR (Decreasing Failure Rate): $h'(t) < 0$ for all t.
- BT (Bathtub): $\exists\, t^*$ such that $h'(t) < 0$ for $t < t^*$ and $h'(t) > 0$ for $t > t^*$.
- UBT (Upside-down Bathtub / Unimodal): $h'(t) > 0$ then < 0.

These classes are preserved under convolution (IFR $*$ IFR = IFR) but not under mixing.

Burn-in policies. If the manufacturer knows the infant-mortality proportion p_1 and the DFR component f_1, it is optimal to burn in components for time t_b^* that minimises total expected cost:

$$C(t_b) = c_b \cdot t_b + c_f \cdot \Pr(\text{field failure after burn-in}).$$

Under a Weibull DFR component, the optimal t_b^* has a closed form when $k < 1$.

Example — Fitting a bathtub hazard to semiconductor failure data

A dataset of $n = 500$ integrated circuits shows 30 early failures in $[0, 100]$ hours, 20 random failures in $(100, 5000]$ hours, and 40 wear-out failures in $(5000, 8000]$ hours. We fit a 3-component Weibull mixture by EM.
Component 1 (burn-in): $\hat{k}_1 = 0.45$, $\hat{\lambda}_1^{-1} = 50$ h, $\hat{p}_1 = 0.08$. **Component 2** (useful life): $\hat{k}_2 = 1.0$, $\hat{\lambda}_2^{-1} = 4000$ h, $\hat{p}_2 = 0.85$. **Component 3** (wear-out): $\hat{k}_3 = 3.2$, $\hat{\lambda}_3^{-1} = 6500$ h, $\hat{p}_3 = 0.07$.
The mixture hazard at $t = 50$ h is dominated by component 1: $h(50) \approx 0.003$ failures/hour. At $t = 3000$ h: $h(3000) \approx 0.00025$ failures/hour (flat). At $t = 7000$ h: $h(7000) \approx 0.0018$ failures/hour (rising).

Competing risks interpretation. The bathtub curve can also arise from three *competing causes of failure*: $T = \min(T_{\text{burn}}, T_{\text{random}}, T_{\text{wear}})$. By the competing risks calculus, the hazard of the minimum is the sum of cause-specific hazards:

$$h(t) = h_{\text{burn}}(t) + h_{\text{random}}(t) + h_{\text{wear}}(t).$$

Each cause is active throughout, but their relative magnitudes shift across time. This additive structure is exact under independence of $T_{\text{burn}}, T_{\text{random}}, T_{\text{wear}}$ — a strong but often workable assumption.

Generalisations of the bathtub model. Several parametric families generate bathtub hazards without mixing:

Modified Weibull extension (Xie–Lai, 1996): $h(t) = \lambda\beta(1+\alpha t)e^{\alpha t}(\lambda t)^{\beta-1}$, combining a power-law with an exponential growth factor for wear-out.

Additive hazard bathtub: $h(t) = at^{-b} + ct^{d}$ (DFR term + IFR term), with $a, b, c, d > 0$. The minimum of h occurs at $t^* = (ab/cd)^{1/(b+d)}$. Simple closed-form cumulative hazard:

$$H(t) = \frac{at^{1-b}}{1-b} + \frac{ct^{1+d}}{1+d}.$$

Optimal burn-in time. Given a bathtub hazard, the burn-in time t_b^* minimises the total cost $C_{\text{total}}(t_b) = c_b(t_b) + c_m \cdot P_f(t_b, t_m)$, where $c_b(t_b)$ is the cost of burning in for time t_b (proportional to t_b plus the expected number of failures during burn-in) and $P_f(t_b, t_m)$ is the probability of field failure during mission time t_m:

$$P_f(t_b, t_m) = \frac{S(t_b) - S(t_b + t_m)}{S(t_b)} = 1 - e^{-[H(t_b+t_m)-H(t_b)]}.$$

Differentiating: $dP_f/dt_b = -h(t_b)P_f(t_b, t_m) + [h(t_b + t_m) - h(t_b)]e^{-[H(t_b+t_m)-H(t_b)]}$.

The optimal t_b^* satisfies $c_b' + c_m dP_f/dt_b = 0$; in the bathtub's decreasing zone $dP_f/dt_b < 0$ (burn-in reduces field failures faster than it costs), so $t_b^* > 0$ as long as the derivative is negative at $t_b = 0$.

Burn-in and warranty cost. Manufacturers offering a warranty of length w minimise expected warranty cost per unit:

$$W(t_b) = c_b \cdot t_b + c_w \cdot M(t_b, t_b + w) \cdot S(t_b)^{-1},$$

where $M(a, b) = \int_a^b h(t)dt$ is the expected number of failures in (a, b) conditional on survival to a. For the additive bathtub model, $M(t_b, t_b + w)$ has closed form.

Example — Burn-in policy for a CMOS memory chip

Bathtub hazard: $h(t) = 2.5 \times 10^{-4}t^{-0.6} + 1.8 \times 10^{-9}t^{1.4}$ (failures/hour, t in hours). Minimum of h at $t^* = (0.6 \times 2.5 \times 10^{-4}/(1.4 \times 1.8 \times 10^{-9}))^{1/2} = (1.5 \times 10^5/2.52 \times 10^{-9})^{0.5} \approx 7740$ h — but this is prohibitively long.
Practical burn-in at $t_b = 168$ h (1 week): $h(168) = 2.5 \times 10^{-4} \times 168^{-0.6} + 1.8 \times 10^{-9} \times 168^{1.4} \approx 1.4\times10^{-5}+4.6\times10^{-7} \approx 1.45\times10^{-5}$ failures/h. The DFR term dominates; burn-in reduces field failure rate from $h(0^+) \to \infty$ to $h(168) = 1.45 \times 10^{-5}$ f/h. Mission field failure probability (1 year = 8760 h) drops from ≈ 0.042 without burn-in to ≈ 0.022 after 168 h burn-in — a 48% reduction.

Environmental stress screening (ESS). A systematic application of burn-in using elevated temperature cycling, vibration, or humidity to precipitate latent defects. The acceleration factor between ESS stress and field stress is given by the Arrhenius or Coffin–Manson models. ESS duration is chosen to eliminate the DFR sub-population while not consuming appreciable life from the main (IFR or useful-life) population.

De-rating and over-stress failure mechanisms. The Stress-Strength interference model (Section 5.1 prelude in Vol. I language): a component fails when applied stress $S_{\text{app}} > S_{\text{strength}}$. If both are random: $P_f = \Pr(S_{\text{app}} > S_{\text{strength}}) = \int_0^\infty F_{\text{strength}}(s) f_{\text{app}}(s)\, ds$. For normal distributions:

$$P_f = \Phi\left(-\frac{\mu_{\text{strength}} - \mu_{\text{app}}}{\sqrt{\sigma_{\text{strength}}^2 + \sigma_{\text{app}}^2}}\right).$$

This is the reliability index $\beta_R = (\mu_{\text{strength}} - \mu_{\text{app}})/\sqrt{\sigma^2_{\text{strength}} + \sigma^2_{\text{app}}}$ of first-order reliability methods (FORM).

Competing mechanisms and the hazard additivity. When K independent physical degradation mechanisms each produce an exponential failure mechanism (constant hazard λ_k while the mechanism is active), the combined hazard from the bathtub model is:

$$h(t) = \sum_{k=1}^{K} \lambda_k(t) \cdot \mathbf{1}[\text{mechanism } k \text{ active at } t].$$

During the burn-in phase, only infant mortality mechanism ($\lambda_1(t)$ DFR) is significant. During useful life: random failures (λ_2 constant). During wear-out: fatigue ($\lambda_3(t)$ IFR). The partition of time zones is empirical.

Component qualification testing and demonstration. To *demonstrate* $R(T) \geq R^*$ with confidence $1 - \alpha$:

- Success-run test (no failures): test $n^* = \ln\alpha / \ln R^*$ units for duration T with zero failures. For $R^* = 0.90$, $\alpha = 0.05$: $n^* = \ln(0.05)/\ln(0.90) = -2.996/-0.1054 \approx 28.4 \to 29$ units.
- One-failure allowed: n^* from the binomial confidence interval $\Pr(\text{Bin}(n^*, 1-R^*) \leq 1) \geq 1-\alpha$.

Example — Zero-failure demonstration testing

Requirement: $R(1000\,\text{h}) \geq 0.95$ with confidence 95%. $n^* = \ln(0.05)/\ln(0.95) = -2.996/-0.0513 \approx 58.4 \to 59$ units tested for 1000 h each with zero failures. Alternatively, test 10 units for $1000 \times \ln(0.05)/\ln(0.95^{10}) \approx 1000 \times 0.503 = 503$ equivalent hours each (time-truncated at 503 h): same statistical confidence with 10 units if failures are independent. Assuming Weibull $k = 1.5$ (IFR), the equivalent time at accelerated stress s_{acc} uses the AFT model: $T_{\text{acc}} = T_{\text{normal}} \cdot (s_{\text{normal}}/s_{\text{acc}})^{\beta}$.

Resampling-based power analysis. Before running a reliability demonstration test, the experimenter can compute the power to detect $R < R^*$ using bootstrap simulation: generate $B = 10^4$ samples of size n^* from an alternative Weibull (with $R(T) = R^* - \delta$ for various $\delta > 0$), count the fraction passing the zero-failure criterion. This gives the operating characteristic (OC) curve of the test plan.

Numerical TTT plot construction. Given ordered failure times $t_{(1)} \leq \cdots \leq t_{(n)}$, compute:

$$H_n(k/n) = \frac{\sum_{i=1}^{k} t_{(i)} + (n-k)t_{(k)}}{\sum_{i=1}^{n} t_{(i)}}, \quad k = 1, \ldots, n.$$

Plot the points $(k/n, H_n(k/n))$ and compare to the 45° diagonal. A graphical test for IFR vs. DFR vs. bathtub requires no parametric assumption.

Confidence bands for the TTT transform. Under exponentiality (H_0), the TTT statistic $H_n(k/n)$ has known distribution (order statistics of a Uniform sample). Barlow–Campo (1975) confidence bands at 95%: $H_n(u) \pm 1.36/\sqrt{n}$ for all u simultaneously (Kolmogorov-type band). Departures outside the band indicate non-exponential hazard.

Example — TTT analysis of motor bearing failures

$n = 15$ bearing failure times (hours): 120, 145, 190, 210, 255, 290, 340, 410, 480, 560, 650, 760, 890, 1050, 1280. Total: $\sum t_{(i)} = 7379$ h.
TTT at $k = 3$ (20th percentile): $H_n(0.20) = (120 + 145 + 190 + 12 \times 190)/7379 = (455+2280)/7379 = 2735/7379 = 0.371$. Since $0.371 > 0.20$ (above diagonal): concave TTT, IFR evidence.
TTT at $k = 8$ (median): $H_n(0.533) = (\sum_{i=1}^{8} t_{(i)} + 7 \times t_{(8)})/7379 = (2240+2870)/7379 = 5110/7379 = 0.693$. Since $0.693 > 0.533$: still above diagonal.
TTT at $k = 13$: $H_n(0.867) = (7379 - t_{(14)} - t_{(15)} + 2t_{(13)})/7379 = (7379 - 1050 - 1280 + 2 \times 890)/7379 = 6829/7379 = 0.925 > 0.867$.
Conclusion: TTT curve consistently above the diagonal throughout $\rightarrow$ **IFR distribution**. Weibull fit confirms: $\hat{k} = 2.3 > 1$.

Bayesian detection of bathtub vs. IFR vs. DFR. Place a mixture prior over the three hazard shape classes:

$$\pi(\text{shape}) = p_{\text{BT}}\delta_{\text{BT}} + p_{\text{IFR}}\delta_{\text{IFR}} + p_{\text{DFR}}\delta_{\text{DFR}}.$$

Compute the marginal likelihood of the data under each class (integrating over the Weibull parameter prior). The posterior probabilities update via Bayes: $\Pr(\text{IFR}|\text{data}) \propto p_{\text{IFR}} \cdot m_{\text{IFR}}(\text{data})$ where $m_{\text{IFR}} = \int L(k, \lambda)\pi(k|k > 1)\pi(\lambda)\, dk\, d\lambda$. This provides a coherent model-uncertainty quantification for the maintenance decision.

5.4 Kaplan–Meier Estimation and Censored Data

Right Censoring and the Kaplan–Meier Estimator

Right censoring. An observation is *right-censored* at c_i if only $\min(T_i, c_i)$ and the indicator $\delta_i = \mathbf{1}[T_i \le c_i]$ are observed. Causes: unit still functioning at study end, lost to follow-up, or competing event.
Kaplan–Meier (KM) estimator. Let $t_1 < t_2 < \cdots < t_m$ be the distinct observed failure times. Let d_j = number of failures at t_j and n_j = number at risk (alive and uncensored) just before t_j. Then:

$$\hat{S}(t) = \prod_{j:\, t_j \le t} \left(1 - \frac{d_j}{n_j}\right).$$

Derivation via counting processes. The KM estimator is the nonparametric MLE of S under right censoring. The likelihood under the discrete hazard model is

$$L = \prod_{j=1}^{m} h_j^{d_j}(1 - h_j)^{n_j - d_j},$$

where $h_j = \Pr(T = t_j \mid T \ge t_j)$. Maximising separately at each t_j: $\hat{h}_j = d_j/n_j$, giving $\hat{S}(t) = \prod_{j:\, t_j \le t}(1 - \hat{h}_j)$. This is exactly the KM estimator.

Nelson–Aalen estimator. The nonparametric MLE of the cumulative hazard is

$$\hat{H}(t) = \sum_{j:\, t_j \le t} \frac{d_j}{n_j},$$

with corresponding survival estimate $\tilde{S}(t) = e^{-\hat{H}(t)}$. The Nelson–Aalen estimate is often preferred in small samples since it avoids the boundary behaviour of KM near 1.

Greenwood's formula for the variance of $\hat{S}(t)$:

$$\widehat{\mathrm{Var}}[\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{j:\, t_j \le t} \frac{d_j}{n_j(n_j - d_j)}.$$

Confidence bands: log-transformed CIs $\hat{S}(t)^{\exp(\pm z_{\alpha/2}\hat{\sigma}(t)/[\hat{S}(t)\ln\hat{S}(t)])}$ preserve $[0,1]$ better than linear CIs.

Example — Kaplan–Meier step function with censoring

Seven bearings are tested; three are removed at inspection without failure (censored, marked $+$):

Time	Event	n_j, d_j, $\hat{S}$
$t=2$	Fail	$n=7$, $d=1$: $\hat{S}(2) = (1-1/7) = 6/7 \approx 0.857$
$t=3^+$	Censored	—
$t=5$	Fail	$n=5$, $d=1$: $\hat{S}(5) = (6/7)(1-1/5) = (6/7)(4/5) \approx 0.686$
$t=6^+$	Censored	—
$t=8$	Fail	$n=3$, $d=1$: $\hat{S}(8) \approx 0.686 \times (2/3) \approx 0.457$
$t=9^+$	Censored	—
$t=11$	Fail	$n=1$, $d=1$: $\hat{S}(11) \approx 0.457 \times 0 = 0$

The step function drops only at uncensored failure times. The censored observations at $t = 3, 6, 9$ are removed from the risk set but contribute information by extending n_j at earlier times.

Log-rank test. To compare two survival curves $\hat{S}_1(t)$ and $\hat{S}_2(t)$ (e.g. treated vs. control), the log-rank statistic is:

$$\chi^2 = \frac{\left(\sum_j w_j(d_{1j} - e_{1j})\right)^2}{\sum_j w_j^2 \hat{V}_j},$$

where $e_{1j} = n_{1j}d_j/n_j$ is the expected failures in group 1 at time t_j, $\hat{V}_j = n_{1j}n_{2j}d_j(n_j - d_j)/(n_j^2(n_j-1))$, and $w_j = 1$ for the unweighted log-rank (Mantel–Haenszel). Under $H_0 : S_1 = S_2$, the statistic is asymptotically χ^2_1.

Type I and Type II censoring. *Type I*: test ends at a fixed time τ; the number of failures D is random. *Type II*: test ends when the r-th failure occurs; the test duration $T_{(r)}$ is random. Type II censoring gives $D = r$ exactly, simplifying the exact conditional inference.

Expected test duration under Type II censoring. For n $\mathrm{Exp}(\lambda)$ units tested to the r-th failure:

$$\mathbb{E}[T_{(r)}] = \frac{1}{\lambda} \sum_{k=n-r+1}^{n} \frac{1}{k} \approx \frac{1}{\lambda} \ln \frac{n}{n-r}.$$

Confidence intervals for survival functions. Beyond Greenwood's formula, two improved CI methods:

Log-log transformation: $\hat{U}(t) = \ln(-\ln \hat{S}(t))$. By the delta method: $\hat{\sigma}_U^2 \approx \frac{\sum_{t_j \le t} d_j/[n_j(n_j-d_j)]}{[\ln \hat{S}(t)]^2}$. CI: $\hat{S}(t)^{\exp(\pm z_{\alpha/2}\hat{\sigma}_U)}$. This is constrained to $(0,1)$ and has superior coverage in small samples.

Bootstrap: resample the (t_i, δ_i) pairs with replacement B times, compute $\hat{S}^{(b)}(t)$ for each resample, use the empirical quantiles as CI. The bootstrap CI automatically accounts for the discrete, step-function nature of $\hat{S}$.

Competing risks in the KM estimator. Under competing risks, the KM estimator *overestimates* the marginal survival of a specific cause (treating other causes as independent censoring), because the latent times are not independent — a unit censored by cause $k \neq j$ need not survive under cause j alone. The correct estimator is the cumulative incidence function (Section 7.5).

Testing for differences in K groups: Weighted log-rank tests. The general weighted log-rank family:

$$Z = \frac{\sum_j w(t_j)(O_{1j} - E_{1j})}{\sqrt{\sum_j w(t_j)^2 \widehat{V}_j}},$$

- $w = 1$: log-rank (equally weights all failure times — best power for PH alternatives).
- $w = \hat{S}(t)$ (Prentice–Wilcoxon): weights early failures more — best power when hazard ratio is large early.
- $w = n_j$: Gehan–Wilcoxon: proportional to risk set size.
- $w = \hat{S}(t)^\rho (1 - \hat{S}(t))^\gamma$ (Fleming–Harrington $G^{\rho,\gamma}$ family): flexible.

Example — Log-rank test for two heat treatment conditions

50 steel pins per group (untreated vs. nitrided). Failure times observed over 2000 hours. Log-rank statistic: $\chi^2 = 12.4$, $p = 0.0004$. Conclusion: strong evidence that nitriding significantly extends fatigue life. Prentice–Wilcoxon statistic: $\chi^2 = 9.8$, $p = 0.0017$ (slightly less extreme than log-rank, consistent with a hazard ratio that increases over time rather than being constant from the start). The KM curves show a clearer separation after $t = 500$ h.

Regression trees for survival. The classification and regression tree (CART) framework extends to survival via the log-rank splitting criterion: at each node, find the split (j, s) on covariate j at threshold s that maximises the log-rank statistic between the two daughter nodes. This gives a nonparametric, interaction-capturing alternative to the Cox model, at the cost of lower statistical efficiency.

Left truncation and delayed entry. In prospective studies, some subjects enter the study only after surviving to time $l_i > 0$ (e.g. a study of disease after age 50: subjects enter the risk set only at age 50). The KM estimator with left truncation:

$$\hat{S}(t) = \prod_{j:\, t_j \leq t} \left(1 - \frac{d_j}{n_j}\right),$$

where n_j now counts only subjects who have both entered (survived to $l_i \leq t_j$) and not yet left (not failed or censored before t_j). The product-limit formula is unchanged, but the risk set n_j correctly excludes subjects not yet entered.

Multiple events per subject: Andersen–Gill model. When subjects can experience repeated events (e.g. recurrent failures of the same machine), the Andersen–Gill counting process model extends Cox:

$$dN_i(t) = Y_i(t) h_0(t) e^{\boldsymbol{\beta}^\top \mathbf{x}_i(t)}\, dt + dM_i(t),$$

where $N_i(t)$ is the number of events by time t, $Y_i(t) = \mathbf{1}[\text{unit } i \text{ is at risk at } t]$, and M_i is a

martingale. The partial likelihood is:

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{\text{all event times}} \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}_{i_t}(t))}{\sum_{k:Y_k(t)=1} \exp(\boldsymbol{\beta}^\top \mathbf{x}_k(t))}.$$

This treats recurrent events as if each failure were from a (possibly different) at-risk unit.

Wei–Lin–Weissfeld (WLW) marginal model. An alternative that models each event number separately with its own partial likelihood, then combines via a robust variance estimator. More flexible than Andersen–Gill but requires specifying a maximum number of events.

Example — Recurrent failures of industrial pumps

$n = 30$ pumps observed over 3 years; total of 127 failure events. Andersen–Gill Cox model with covariate x_1 = pump age (years). $\hat{\beta}_1 = 0.18$ ($p = 0.002$): each additional year of pump age increases the instantaneous failure rate by 20%. Robust standard errors (Lin–Wei sandwich estimator) are 35% larger than the naive Cox SEs, reflecting within-unit correlation between successive failures. The marginal model (WLW) gives consistent but less efficient estimates.

Nonparametric tests for recurrent events. The *LWYY test* (Lin–Wei–Yang–Ying) compares two groups for equality of mean cumulative failure functions (MCF): $\mu_k(t) = \mathbb{E}[N_k(t)]$. Test statistic: $Z = \int w(t)[\hat{\mu}_1(t) - \hat{\mu}_2(t)]\, dt/\hat{\sigma}_Z$. Under $H_0 : \mu_1 = \mu_2$, $Z \to \mathcal{N}(0,1)$. The MCF estimator: $\hat{\mu}(t) = \int_0^t \hat{R}^{-1}(s)\, dN(s)$ where $\hat{R}(s) = n^{-1}\sum_i Y_i(s)$ and $N(s) = \sum_i N_i(s)$.

5.5 Proportional Hazards and the Cox Model

Cox Proportional Hazards Model

The Cox (1972) model specifies:

$$h(t \mid \mathbf{x}) = h_0(t) \cdot \exp(\boldsymbol{\beta}^\top \mathbf{x}),$$

where $h_0(t)$ is an unspecified *baseline hazard*, $\mathbf{x} \in \mathbb{R}^p$ is a covariate vector, and $\boldsymbol{\beta} \in \mathbb{R}^p$ are regression coefficients. The model is *semiparametric*: h_0 is nonparametric, $\boldsymbol{\beta}$ is estimated from the data.

Partial likelihood. Cox observed that $\boldsymbol{\beta}$ can be estimated without specifying h_0 via the *partial likelihood*:

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{j=1}^{m} \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}_{(j)})}{\sum_{i \in \mathcal{R}(t_j)} \exp(\boldsymbol{\beta}^\top \mathbf{x}_i)},$$

where $\mathbf{x}_{(j)}$ is the covariate vector of the unit failing at t_j, and $\mathcal{R}(t_j) = \{i : T_i \geq t_j\}$ is the risk set at t_j. This is a product of conditional probabilities: given that one failure occurs at t_j, what is the probability it was unit (j)?

Score equations and Fisher information. The log-partial-likelihood score is:

$$U(\boldsymbol{\beta}) = \sum_{j=1}^{m} \left[\mathbf{x}_{(j)} - \bar{\mathbf{x}}(\boldsymbol{\beta}, t_j)\right], \quad \bar{\mathbf{x}}(\boldsymbol{\beta}, t) = \frac{\sum_{i \in \mathcal{R}(t)} \exp(\boldsymbol{\beta}^\top \mathbf{x}_i)\mathbf{x}_i}{\sum_{i \in \mathcal{R}(t)} \exp(\boldsymbol{\beta}^\top \mathbf{x}_i)}.$$

$U(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ gives the MLE. The observed information $I(\boldsymbol{\beta}) = -\partial U/\partial \boldsymbol{\beta}$ yields standard errors and Wald tests.

Breslow estimator of $H_0(t)$. Given $\hat{\boldsymbol{\beta}}$, the cumulative baseline hazard is estimated by:

$$\hat{H}_0(t) = \sum_{j:\, t_j \le t} \frac{d_j}{\sum_{i \in \mathcal{R}(t_j)} \exp(\hat{\boldsymbol{\beta}}^\top \mathbf{x}_i)}.$$

Proportionality assumption and diagnostics. The key assumption $h_i(t)/h_k(t) = \exp(\boldsymbol{\beta}^\top(\mathbf{x}_i - \mathbf{x}_k))$ is *constant in time*. Diagnostics:

- *Schoenfeld residuals*: plot $\hat{\beta}_l + \hat{r}_{lj}$ against t_j; a trend indicates non-proportionality.
- *Log–log plot*: $\ln(-\ln \hat{S}_k(t))$ vs. $\ln t$ should be parallel for different covariate groups.
- *Time-varying coefficients*: extend to $h(t \mid \mathbf{x}) = h_0(t)\exp(\boldsymbol{\beta}(t)^\top \mathbf{x})$, estimated via B-splines or piecewise constants.

Example — Cox model for turbine bearing failures

A fleet of $n = 80$ turbines is monitored for bearing failures over 3 years. Covariates: $x_1 =$ operating temperature (standardised), $x_2 =$ bearing load index. Partial likelihood MLE: $\hat{\beta}_1 = 0.62$ (95% CI: $[0.31, 0.93]$), $\hat{\beta}_2 = 0.38$ (95% CI: $[0.11, 0.65]$). Hazard ratios: $e^{\hat{\beta}_1} = 1.86$ per SD increase in temperature — bearings in hotter turbines fail at 86% higher rate at any given time. $e^{\hat{\beta}_2} = 1.46$ per unit load increase. Schoenfeld residuals show no temporal trend (proportionality holds). Predicted survival at 2 years for a high-temperature, high-load turbine ($\mathbf{x} = (1.5, 2.0)$): $\hat{S}(2 \mid \mathbf{x}) = [\hat{S}_0(2)]^{\exp(0.62\times 1.5 + 0.38 \times 2.0)} = 0.71^{3.83} \approx 0.26$.

Extensions. (i) *Stratified Cox*: separate $h_0^{(s)}(t)$ per stratum s, shared $\boldsymbol{\beta}$. (ii) *Frailty models*: add a random effect Z_i (gamma or log-normal) to account for unobserved heterogeneity: $h(t \mid \mathbf{x}_i, Z_i) = Z_i h_0(t) e^{\boldsymbol{\beta}^\top \mathbf{x}_i}$. (iii) *Competing risks*: cause-specific Cox models, one per cause c, with all other causes treated as censoring.

Part 6

Fatigue, Fracture, and Degradation Mechanics

Mathematical models of cumulative damage in materials: S-N curves, Wöhler's law, fracture mechanics, and stochastic degradation processes.

Additive hazard models. The Lin–Ying additive hazard model specifies:

$$h(t \mid \mathbf{x}) = h_0(t) + \boldsymbol{\beta}(t)^\top \mathbf{x},$$

where $\boldsymbol{\beta}(t)$ are time-varying coefficients representing the *absolute* (not relative) change in hazard. Unlike PH, covariates shift hazard by a constant amount; this is appropriate when the covariate effect is additive on the hazard scale rather than multiplicative. Estimation via least squares on martingale integrals.

Frailty models. A *frailty* $Z_i > 0$ is a random effect representing unobserved heterogeneity: $h(t \mid \mathbf{x}_i, Z_i) = Z_i h_0(t) e^{\boldsymbol{\beta}^\top \mathbf{x}_i}$. Conditional on Z, failures are independent. The marginal survival is:

$$S(t \mid \mathbf{x}) = \mathbb{E}_Z[e^{-Z H_0(t) e^{\boldsymbol{\beta}^\top \mathbf{x}}}] = \mathcal{L}_Z(H_0(t) e^{\boldsymbol{\beta}^\top \mathbf{x}}),$$

where $\mathcal{L}_Z(s) = \mathbb{E}[e^{-sZ}]$ is the Laplace transform of the frailty distribution. For $Z \sim \text{Gamma}(1/\theta, 1/\theta)$ (mean 1, variance θ):

$$S(t \mid \mathbf{x}) = \left(1 + \theta H_0(t) e^{\boldsymbol{\beta}^\top \mathbf{x}}\right)^{-1/\theta},$$

yielding the *proportional hazard frailty* or Burr survival model. The marginal hazard is no longer proportional.

Penalised Partial Likelihood

For high-dimensional covariates ($p \gg n$) in survival analysis, regularise the Cox log-partial-likelihood:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \left[\mathcal{L}(\boldsymbol{\beta}) - \lambda_1 \|\boldsymbol{\beta}\|_1 - \lambda_2 \|\boldsymbol{\beta}\|_2^2 \right],$$

where the L^1 term (LASSO) induces sparsity and the L^2 term (ridge) handles correlated covariates. The elastic net penalty (λ_1, λ_2) is tuned by cross-validated partial likelihood. The coordinate descent algorithm cycles through coordinates, applying the soft-thresholding operator at each step: $\hat{\beta}_j \leftarrow \mathcal{S}(\tilde{\beta}_j, \lambda_1/\tilde{I}_{jj})/(1 + \lambda_2/\tilde{I}_{jj})$.

Deep learning for survival. DeepSurv (Katzman et al., 2018) replaces the linear predictor $\boldsymbol{\beta}^\top \mathbf{x}$ in Cox with a neural network $f_\theta(\mathbf{x})$: $h(t \mid \mathbf{x}) = h_0(t) \exp(f_\theta(\mathbf{x}))$. Training minimises the negative partial likelihood. DeepHit (Lee et al., 2018) directly models the joint distribution of (T, C) via a multi-task softmax output, accommodating non-proportional hazards and competing risks. These models sacrifice interpretability for predictive performance, with SHAP values used for post-hoc explanation.

Time-varying covariates. If $\mathbf{x}_i(t)$ evolves over time (e.g. cumulative drug dose, vibration level), the Cox model extends to:

$$h(t \mid \mathcal{F}_{t^-}) = h_0(t) \exp(\boldsymbol{\beta}^\top \mathbf{x}_i(t)),$$

where $\mathcal{F}_{t^-}$ is the filtration up to time t^-. Estimation uses the counting process formulation: the partial likelihood numerator uses $\mathbf{x}_i(t_j)$ (covariate value at the failure time), and the risk set denominator uses $\mathbf{x}_k(t_j)$ for all $k \in \mathcal{R}(t_j)$.

Example — Joint modelling of degradation and failure

A turbine blade's degradation trajectory $X_i(t)$ (measured vibration) is modelled jointly with the failure indicator. The joint model:

$$X_i(t) = \mu_i + b_i t + \sigma_\epsilon \epsilon_i(t) \quad \text{(LME longitudinal sub-model)},$$
$$h_i(t) = h_0(t) \exp(\alpha_1 X_i(t) + \alpha_2 \dot{X}_i(t) + \boldsymbol{\beta}^\top \mathbf{w}_i) \quad \text{(survival sub-model)}.$$

The current value $X_i(t)$ and slope $\dot{X}_i(t)$ enter the hazard as time-varying covariates. Joint MLE via EM: E-step computes $\mathbb{E}[X_i(t) \mid T_i, \delta_i, \hat{\theta}]$ using Kalman smoothing; M-step updates $(\hat{\mu}, \hat{b}, \hat{\sigma}_\epsilon)$ and $(\hat{h}_0, \hat{\alpha}, \hat{\boldsymbol{\beta}})$ simultaneously. This outperforms landmark analysis and two-stage methods by propagating uncertainty from the degradation model to the survival model.

Causal inference in survival analysis. Observational data suffer from confounding: unhealthy units receive more aggressive maintenance (treatment-by-indication bias). Causal hazard ratio estimation uses:

- *Propensity score weighting*: weight unit i by $1/\hat{e}(\mathbf{x}_i)$ (treated) or $1/(1 - \hat{e}(\mathbf{x}_i))$ (control), where $\hat{e}(\mathbf{x}) = \Pr(\text{treated} \mid \mathbf{x})$ is the propensity score. The weighted KM curves estimate counterfactual survival.

- *Instrumental variable* (*IV*): use a variable Z (correlated with treatment but not failure except through treatment) for 2-stage estimation. In reliability: Z = supplier batch (affects component quality but not operating conditions).
- *G-estimation of structural nested accelerated failure time models*: directly estimates the causal AFT parameter ψ such that $T^{(0)} = T^{(\text{obs})} e^{-\psi A}$ (where A = treatment indicator) has the same distribution as the no-treatment counterfactual.

Exact tests for small samples. When n is small (≤ 20) and the asymptotic log-rank distribution is unreliable, exact permutation tests are available: under H_0, randomly re-assign the failure/censoring labels among all units and recompute the log-rank statistic. The p-value is the fraction of B permutations with statistic $\geq$ observed. For $K = 2$ groups and small samples, the exact conditional distribution of D_{1j} given n_{1j}, n_j, d_j is hypergeometric (the Mantel test).

Bayesian survival analysis. A fully Bayesian approach places priors on the hazard function and obtains posterior distributions:

- *Beta process prior*: $H_0(\cdot) \sim \text{BP}(c, H^*)$, where H^* is the prior mean cumulative hazard. The posterior given failure data is also a Beta process (conjugate), and the posterior mean hazard increment at t_j is a shrinkage estimator between the prior $dH^*(t_j)$ and the Nelson–Aalen increment.
- *Cox with Gaussian process prior*: $\ln h_0(t) \sim \text{GP}(m(t), k(t, t'))$. Posterior inference via MCMC (slice sampling or HMC).

Example — Bayesian Cox model for transformer failure prediction

A power utility has $n = 45$ transformers with 8 failures over 20 years. Prior: exponential baseline hazard with rate $\lambda_0 \sim \text{Gamma}(2, 0.1)$ (mean $h_0 = 0.02$/year). Covariates: x_1 = voltage rating (kV, standardised), x_2 = load factor. Prior on $\boldsymbol{\beta}$: $\mathcal{N}(\mathbf{0}, 2\mathbf{I})$ (weakly informative).
Posterior (4 chains $\times$ 2000 HMC samples): $\hat{\beta}_1 = 0.41$ (95% credible interval $[0.08, 0.74]$) — higher voltage transformers fail faster. $\hat{\beta}_2 = 0.29$ ($[-0.12, 0.70]$) — load factor effect uncertain. Posterior predictive failure probability at 10 years for an average transformer: $\hat{P}(T \leq 10) = 0.18$ (90% CI: $[0.11, 0.26]$).

Survival forest and ensemble methods. The random survival forest (Ishwaran–Kogalur, 2008) grows an ensemble of survival trees, each using a random subset of features and a random training subset. Node splitting uses the log-rank statistic. The ensemble's cumulative hazard is $\hat{H}(t \mid \mathbf{x}) = B^{-1} \sum_{b=1}^{B} \hat{H}_b(t \mid \mathbf{x})$. Variable importance is computed via permutation: $I_j = \widehat{\text{error}}_{j,\text{permuted}} - \widehat{\text{error}}_{\text{original}}$, measuring the increase in prediction error when feature j is randomly permuted.

6.1 The Wöhler (S-N) Curve and Basquin's Law

S-N Relationship and Basquin's Law

The *Wöhler curve* (or S-N curve) empirically relates the stress amplitude S (in MPa) applied to a specimen to the number of cycles N to failure. **Basquin's Law** (1910) states:

$$S^m \cdot N = C, \qquad \text{equivalently} \quad N = C \cdot S^{-m},$$

where $m > 0$ (Basquin exponent, typically $3 \leq m \leq 8$ for metals) and $C > 0$ are material constants. In log-space: $\ln N = \ln C - m \ln S$.

Derivation and physical basis. Under cyclic loading, plastic deformation at microscopic crack tips accumulates. Energy dissipated per cycle scales as $\Delta W \propto S^m$, and failure occurs when cumulative damage $D = N \cdot \Delta W$ reaches a critical threshold D_c: giving $N = D_c / \Delta W \propto S^{-m}$.

Endurance limit. Many ferrous metals exhibit a *fatigue limit* (or endurance limit) S_e: below this stress amplitude, fatigue failure does not occur in a finite number of cycles. The S-N curve becomes horizontal for $S < S_e$. Non-ferrous metals (aluminium, copper) typically have no true endurance limit — the S-N curve continues to decrease without levelling off, so an "endurance strength" at 10^7 or 10^8 cycles is used instead.

Statistical scatter. For a fixed S, $\ln N$ is approximately normally distributed. Thus:

$$\ln N \sim \mathcal{N}(\ln C - m \ln S,\ \sigma_\epsilon^2),$$

and the P-S-N (probability-stress-life) curve gives N_p for survival probability p:

$$N_p(S) = \exp\!\big(\ln C - m \ln S + z_p \sigma_\epsilon\big),$$

where $z_p = \Phi^{-1}(p)$. In design, $N_{0.01}$ (1st percentile) is used for safety-critical components.

Staircase method. For estimation of S_e: run specimens alternately above and below the estimated endurance limit; the MLE of S_e and its standard deviation come from the pattern of failures and run-outs. Dixon and Massey (1951) show the estimator is asymptotically efficient.

Example — Basquin fit for a steel alloy

Test data at three stress levels for AISI 4340 steel:

S (MPa)	N (cycles)
700	3.2×10^4
500	4.1×10^5
400	3.0×10^6

Log-space regression: $\ln N = a - m \ln S$. Using the first and third points:

$$m = \frac{\ln(3.0 \times 10^6) - \ln(3.2 \times 10^4)}{\ln 700 - \ln 400} = \frac{4.541}{0.560} \approx 8.11.$$

Then $\ln C = \ln N + m \ln S = \ln(3.2 \times 10^4) + 8.11 \ln 700 = 10.37 + 53.93 = 64.30$, giving $C = e^{64.30} \approx 8.3 \times 10^{27}$. At $S = 450$ MPa: $N = 8.3 \times 10^{27} \cdot 450^{-8.11} \approx 8.0 \times 10^5$ cycles.

Temperature and environment effects. In the Coffin–Manson relation for low-cycle fatigue ($N < 10^4$), plastic strain amplitude $\Delta\varepsilon_p/2$ replaces stress:

$$\frac{\Delta\varepsilon_p}{2} = \varepsilon_f'(2N)^c,$$

where ε_f' is the fatigue ductility coefficient and $c \in [-0.5, -0.7]$ is the Coffin–Manson exponent. Elevated temperature reduces S_e and ε_f' while increasing c; this is modelled via an Arrhenius factor $\exp(-Q/RT)$ on the relevant material constants.

Multiaxial fatigue. Real components experience combined bending, torsion, and tension. The von Mises stress amplitude $\sigma_{\text{VM}} = \sqrt{\sigma_a^2 + 3\tau_a^2}$ (for proportional loading) is used as

the equivalent stress in the S-N curve. For non-proportional loading (phase shift between σ and τ), the critical plane method (Fatemi–Socie, Smith–Watson–Topper) identifies the plane maximising a damage parameter combining normal and shear strains.

Mean stress correction. For $\sigma_{\min} \neq -\sigma_{\max}$ (non-zero mean stress $\sigma_m = (\sigma_{\max} + \sigma_{\min})/2$):

- **Goodman**: $S_a/S_e + \sigma_m/\sigma_{UTS} = 1$ (conservative, common in design).
- **Gerber**: $S_a/S_e + (\sigma_m/\sigma_{UTS})^2 = 1$ (better fit to data at low mean stress).
- **Soderberg**: $S_a/S_e + \sigma_m/\sigma_Y = 1$ (most conservative; uses yield strength).
- **ASME Elliptic**: $(S_a/S_e)^2 + (\sigma_m/\sigma_Y)^2 = 1$.

The modified Goodman diagram plots S_a vs. σ_m with these curves as failure boundaries; design points must lie inside all applicable curves.

Stress Concentration Factor and Notch Effect

For a component with a geometric notch (hole, fillet, groove), the theoretical stress concentration factor $K_t = \sigma_{\max}/\sigma_{\text{nom}}$ amplifies the nominal stress. The *fatigue notch factor* $K_f \leq K_t$ accounts for the notch sensitivity:

$$K_f = 1 + \frac{K_t - 1}{1 + a/r},$$

where r is the notch root radius and a is the material's *Neuber constant* (order of microns for steels, larger for aluminium). For very sharp notches ($r \to 0$), $K_f \to K_t$ but the actual fatigue strength is higher than K_t alone predicts because the high-stress region is confined to a sub-grain volume.

Size effect in fatigue. Laboratory S-N data are derived from small smooth specimens. For large components:

$$S_e^{\text{comp}} = S_e^{\text{lab}} \cdot C_{\text{size}} \cdot C_{\text{surface}} \cdot C_{\text{load}} \cdot C_{\text{temp}} \cdot C_{\text{reliability}},$$

where typical Marin correction factors: $C_{\text{size}} \approx 0.75$ for $d > 50$ mm; C_{surface} depends on finish (machined ≈ 0.9; hot-rolled ≈ 0.6; corroded < 0.4); $C_{\text{reliability}}$ for 99% reliability ≈ 0.75.

Probabilistic S-N approach and p-S-N diagram. Considering test scatter:

$$\ln N = \mu_{\ln N}(S) - k_p \sigma_{\ln N},$$

where $k_p = \Phi^{-1}(1-p)$ is the reliability factor (e.g. $k_p = 2.326$ for $p = 0.99$). The vertical scatter in $\ln N$ at fixed S is $\sigma_{\ln N}$, typically 0.2–0.5. The probabilistic S-N band requires testing at multiple stress levels with replications, estimating $\mu_{\ln N}(S)$ and $\sigma_{\ln N}(S)$ simultaneously.

Example — p-S-N curves for titanium alloy Ti-6Al-4V

Tests at $S \in \{600, 500, 450\}$ MPa with $n = 8$ replicates per level.

S (MPa)	$\overline{\ln N}$	$s_{\ln N}$	$N_{50\%}$
600	11.12	0.28	6.7×10^4
500	12.81	0.31	3.6×10^5
450	13.94	0.35	1.1×10^6

OLS on $(\ln S, \overline{\ln N})$: slope $\hat{m} = -9.4$, intercept $\widehat{\ln C} = 74.1$. 99% reliability curve:

$\ln N_{0.01}(S) = \ln\hat{C} - \hat{m}\ln S - 2.326 \times 0.31 = 74.1 - 9.4\ln S - 0.72$. At $S = 500$ MPa: $N_{0.01} = \exp(12.81 - 0.72) = \exp(12.09) = 1.77 \times 10^5$ cycles — less than half the median life.

Variable amplitude loading and cycle counting. For a random load spectrum $\sigma(t)$, the rainflow method extracts cycle amplitudes $S_1, S_2, \ldots, S_K$ and mean stresses $\sigma_{m1}, \ldots, \sigma_{mK}$. After applying mean stress correction (Goodman), the equivalent damage per block is:

$$D_{\text{block}} = \sum_{i=1}^{K} \frac{n_i}{N(S_i^{\text{corrected}})}.$$

The *blocks to failure* is $N_{\text{blocks}} = 1/D_{\text{block}}$, giving total cycles to failure.

Thermomechanical fatigue (TMF). Components subject to cyclic thermal gradients (gas turbine blades, engine pistons) undergo both mechanical and thermal cycling simultaneously. The TMF life is governed by the phase angle ϕ between temperature and strain cycles: in-phase TMF ($\phi = 0°$, heating under tensile strain) accelerates oxidation-fatigue interaction; out-of-phase TMF ($\phi = 180°$, cooling under tensile strain) promotes fatigue-creep interaction. The isothermal Basquin/Coffin-Manson laws must be modified by damage mechanics terms.

Fretting fatigue. At press-fitted joints (e.g. dovetail roots of turbine blades), micro-slip at the contact surface introduces additional crack initiation sites. The fretting fatigue life is reduced by a factor $k_f^{\text{fret}} = 1 - (1-R)(\sigma_c/\sigma_y)$ relative to plain fatigue, where σ_c is the contact normal stress. The combined stress intensity at the fretting scar: $\Delta K_{\text{eff}} = \sqrt{\Delta K_I^2 + \Delta K_{II}^2}$ accounts for mixed-mode loading.

Short crack behaviour. Cracks shorter than the grain size or the plastic zone radius violate LEFM assumptions. El Haddad's short crack correction modifies the threshold:

$$\Delta K_{\text{th}}(a) = \Delta\sigma_e\sqrt{\pi(a + a_0)}, \quad a_0 = \frac{1}{\pi}\left(\frac{\Delta K_{\text{th,lc}}}{\Delta\sigma_e}\right)^2,$$

where a_0 is the El Haddad length (a material constant, typically 10–100 μm). For $a \ll a_0$, the effective threshold $\Delta K_{\text{th}} \approx \Delta\sigma_e\sqrt{\pi a_0}$ is constant (non-propagating crack condition). This explains why cracks below the threshold do not grow even under nominally damaging loading.

Example — Kitagawa–Takahashi diagram for 4340 steel

Endurance limit: $\Delta\sigma_e = 550$ MPa. Threshold SIF: $\Delta K_{\text{th}} = 8$ MPa$\sqrt{\text{m}}$. El Haddad length: $a_0 = (8/(550\sqrt{\pi}))^2/\pi = (0.00822)^2/\pi = 2.15 \times 10^{-5}$ m $= 21.5\,\mu$m.
The Kitagawa–Takahashi diagram plots $\Delta\sigma_{\text{threshold}}$ vs. a: for $a > a_0$, $\Delta\sigma_{\text{th}} = \Delta K_{\text{th}}/(Y\sqrt{\pi a})$ (LEFM regime, decreasing); for $a < a_0$, $\Delta\sigma_{\text{th}} = \Delta\sigma_e$ (plain fatigue, constant). The transition at $a_0 = 21.5\,\mu$m defines the critical crack size: cracks smaller than 21.5 μm are non-damaging at $\Delta\sigma = 550$ MPa.

Cumulative damage under spectrum loading: stochastic approach. Random loading $\sigma(t)$ is modelled as a stationary Gaussian process with power spectral density $G(\omega)$. The expected damage rate per unit time (Dirlik formula):

$$E[\dot{D}] = \frac{\nu_0}{C}\left[\frac{D_1 Q}{M_1}e^{-Z_1/Q} + \frac{D_2 Z}{M_2^2}e^{-Z^2/(2M_2^2)} + D_3 Z e^{-Z^2/2}\right]^{-1/m} \cdot (\text{Dirlik integral}),$$

where ν_0 is the zero-crossing rate, M_1, M_2 are spectral moments, and D_1, D_2, D_3, Q, Z are Dirlik coefficients (functions of the spectral width ε). This gives fatigue damage in closed form from the load PSD without explicit cycle counting — essential for structural dynamics and vibration fatigue.

6.2 Linear and Nonlinear Cumulative Damage: Miner's Rule and Beyond

Palmgren–Miner Linear Damage Rule

Under variable-amplitude loading with k distinct stress levels $S_1, \dots, S_k$, applied for n_i cycles each, the *Miner's rule* postulates that failure occurs when the *damage index* reaches unity:

$$D = \sum_{i=1}^{k} \frac{n_i}{N_i(S_i)} \geq 1,$$

where $N_i(S_i)$ is the number of cycles to failure at constant amplitude S_i (from the S-N curve). Failure is predicted at $D = 1$ (Miner's critical damage).

Derivation. Assuming damage accumulates linearly at rate $1/N_i$ per cycle under stress S_i, the total damage after n_i cycles at each level is the sum above. The linearity assumption is exact if each stress level independently consumes a fraction n_i/N_i of the material's fatigue life.

Known deficiencies. Experimental data shows that the actual critical damage D_c at failure varies from 0.3 to 3.0 depending on the loading sequence:

- High-to-low loading ($S_1 > S_2 > \cdots$): $D_c < 1$ (Miner is non-conservative — dangerous).
- Low-to-high loading: $D_c > 1$ (Miner is conservative).

The loading sequence effect arises from residual stresses and crack retardation/acceleration.

Nonlinear cumulative damage models. Several refinements exist:

- **Marco–Starkey** (1954): $D = \sum_i (n_i/N_i)^{p(S_i)}$ with $p(S) > 1$ for high stress.
- **Corten–Dolan**: $D = \sum_i n_i r_i d_0^{-1}$ where $r_i = (S_i/S_{\max})^m$ weights each level by relative severity.
- **Damage curve approach** (Leve, 1969): the damage α after n cycles at S is $\alpha = (n/N(S))^{q(\alpha)}$, a nonlinear ODE in n.

Rainflow counting. For irregular time-series loads $\sigma(t)$, the Rainflow algorithm (Matsuishi–Endo, 1968) extracts a set of closed stress cycles by identifying peaks, troughs, and their pairing. The algorithm processes the load history as a series of "half-cycles" collected in stacks, pairing each rise with the nearest fall of equal amplitude. The output is a histogram n_i over stress ranges $\Delta\sigma_i$, which feeds directly into Miner's sum.

Example — Miner's rule under two-level loading

A shaft is subjected to: $n_1 = 10^4$ cycles at $S_1 = 600$ MPa ($N_1 = 3 \times 10^4$), then $n_2 =$

2×10^5 cycles at $S_2 = 400$ MPa ($N_2 = 1.5 \times 10^6$). Miner damage:

$$D = \frac{10^4}{3 \times 10^4} + \frac{2 \times 10^5}{1.5 \times 10^6} = 0.333 + 0.133 = 0.467.$$

Remaining life at $S_2 = 400$ MPa: $n_{\text{rem}} = (1 - 0.467) \times 1.5 \times 10^6 \approx 8.0 \times 10^5$ additional cycles.
Note: since high stress was applied first, experimental evidence suggests the actual remaining life may be *shorter* than Miner predicts (observed $D_c \approx 0.7$ for this loading sequence), so the safe design should use $D_c = 0.7$ as the failure criterion.

Probabilistic Miner's rule. When the S-N curve has scatter, $N_i(S_i)$ is random with log-normal distribution. Then $D = \sum_i n_i/N_i$ is a sum of log-normal variables, and $\Pr(D \geq D_c)$ can be computed by Monte Carlo or by the first-order second-moment (FOSM) approximation:

$$\mathbb{E}[D] = \sum_i n_i \mathbb{E}[N_i^{-1}], \qquad \operatorname{Var}[D] = \sum_i n_i^2 \operatorname{Var}[N_i^{-1}].$$

Under log-normality, $\mathbb{E}[N^{-1}] = e^{-\mu_{\ln N} + \sigma^2_{\ln N}/2}$.

Sequence effects and interaction damage models. The Miner rule's inability to capture sequence effects is addressed by the *damage curve approach* (DCA). Define a *damage function* $D(n, N_i)$ as the actual damage after n cycles at stress S_i. DCA postulates:

$$D(n, N_i) = \left(\frac{n}{N_i} \right)^{q_i},$$

where $q_i = q(S_i) > 1$ for high stress (damage accumulates faster than linearly) and $q_i < 1$ for low stress. Sequence effects enter because high-stress damage at $q_{\text{high}} > 1$ shifts the starting point on the $q_{\text{low}} < 1$ curve, consuming more life than Miner predicts.

Bilinear damage rule. Chaboche (1988) proposed a two-phase damage:

$$D(N) = \begin{cases} 1 - (1 - D_0)^{1/(1-\alpha)} & \text{phase I (crack initiation)}, \\ 1 - \left[1 - (1 - D_0)^{1/(1+\beta)}\right]^{1/(1+\beta)} & \text{phase II (crack propagation)}, \end{cases}$$

where D_0 is the transition damage. The bilinear form better captures the distinct microstructural mechanisms of initiation (dislocation band formation) and propagation (microcrack coalescence).

Energy-based damage approaches. The total plastic energy dissipated per cycle $\Delta W_p = \oint \sigma \, d\varepsilon_p$ is proportional to the area of the hysteresis loop. Morrow's criterion: failure when $W_p^{\text{cum}} = \int_0^{N_f} \Delta W_p \, dN = W_c$ (critical energy). For a cyclic hardening material, ΔW_p evolves with N; the cumulative energy criterion integrates the Ramberg–Osgood cyclic curve.

Example — Rainflow + Miner for a wind turbine blade

Load spectrum from 1-hour anemometry: 847 cycles extracted by rainflow. Grouped into 5 amplitude bins:

$\Delta\sigma_i$ (MPa)	n_i (cycles/h)	N_i (Basquin, $m=5$)	$d_i = n_i/N_i$
80	420	3.5×10^8	1.2×10^{-6}
100	215	9.1×10^7	2.4×10^{-6}
130	98	1.7×10^7	5.8×10^{-6}
160	42	3.8×10^6	1.1×10^{-5}
200	12	6.7×10^5	1.8×10^{-5}

Hourly damage: $D_{\text{hour}} = \sum d_i = 3.85 \times 10^{-5}$. Design life to Miner: $N_f = 1/D_{\text{hour}} = 26000$ hours ≈ 3.0 years of continuous operation. With Miner's critical damage $D_c = 0.7$ (conservative): $N_f^{\text{safe}} = 18200$ h ≈ 2.1 years.

Spectrum loading and load clipping. High-amplitude cycles are rare but dominate Miner damage ($d_i \propto S^m$ with $m = 5$). Load *clipping* at S_{clip} truncates cycles above a threshold, reducing test time at a cost of non-conservatism. The truncation error in Miner damage is:

$$\Delta D = \sum_{i:\, S_i > S_{\text{clip}}} \frac{n_i}{N(S_i)},$$

bounded by $\Delta D < n_{\text{total}} \cdot N(S_{\text{clip}})^{-1}$. Standardised spectra (FALSTAFF for fighters, TWIST for transport wings, WISPER for wind turbines) enable inter-laboratory comparisons.

Fatigue reliability using Monte Carlo. Given uncertain material parameters $(\ln C, m) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and stochastic loading spectrum (random n_i), generate $N_{\text{MC}} = 10^5$ damage samples:

$$D^{(k)} = \sum_i n_i^{(k)} \left[C^{(k)} S_i^{m^{(k)}}\right]^{-1}.$$

The failure probability $P_f = \Pr(D \geq D_c) = \#\{D^{(k)} \geq D_c\}/N_{\text{MC}}$ and its confidence interval follow directly. Variance decomposition (Sobol' indices) identifies the dominant source of variability: typically m (Basquin exponent) contributes 60–80% of the variance in $\ln D$.

Creep-fatigue interaction. At elevated temperatures, cyclic loading combines fatigue damage $D_f = \sum n_i/N_i$ with creep damage $D_c = \sum t_k/t_r(T_k, \sigma_k)$, where t_r is the creep rupture time at temperature T_k and stress σ_k (from the Larson–Miller parameter). The ASME NH code uses an *interaction diagram*: failure when $D_f + D_c \geq 1$ (linear interaction line) or when the (D_f, D_c) point lies outside the non-linear interaction envelope (bilinear or MPC Omega model).

Continuum damage mechanics (CDM). The Kachanov–Rabotnov approach introduces a scalar damage variable $\omega \in [0, 1]$ ($\omega = 0$: undamaged; $\omega = 1$: failed) with evolution law:

$$\dot{\omega} = \frac{A\sigma^r}{(1-\omega)^k},$$

where A, r, k are material constants. Integrating: $\omega(t) = 1 - [1 - (k+1)A\sigma^r t]^{1/(k+1)}$, and failure at $\omega = 1$: $t_f = 1/[(k+1)A\sigma^r]$. For cyclic loading, CDM is extended by replacing t with N cycles: $d\omega/dN = f(\sigma, \omega)$.

Fatigue crack growth threshold and closure. The threshold ΔK_{th} is not a material constant: it depends on $R = K_{\text{min}}/K_{\text{max}}$, environment (vacuum vs. air vs. corrosive), and crack size (El Haddad). The NASGRO threshold function:

$$\Delta K_{\text{th}}(R) = \Delta K_{\text{th},0} \left(\frac{1-f}{1-f_0}\right)^{1/(1-C_{\text{th}}R)},$$

where f is the crack opening function (from the Newman plasticity-induced closure model) and C_{th} is a material constant.

Example — Creep-fatigue interaction in steam turbine rotor

Rotor operates at 600°C with daily start-stop cycles ($n_f = 365$/year) and sustained load periods. Fatigue damage per year: $D_f = 365/N_f(600°C)$. At 600°C and $\sigma = 200$ MPa, $N_f = 1.5 \times 10^4$ cycles (isothermal LCF): $D_f = 365/15000 = 0.0243$/year. Creep damage: operating 8000 h/year at $T = 600$°C, $\sigma = 200$ MPa; Larson–Miller gives $t_r = 6.5 \times 10^4$ h: $D_c = 8000/65000 = 0.123$/year. Total: $D_f + D_c = 0.147$/year. Design life: ~ 6.8 years to interaction envelope crossing. With a safety factor of 2: allow ≤ 3.4 years between major overhauls.

Block cycle loading and equivalent damage. In a standardised loading spectrum with B blocks of n_i cycles at S_i, the total damage after m blocks is $D = m \cdot D_{\text{block}}$ (linear). The *equivalent constant amplitude* stress that produces the same damage in the same number of cycles: $S_{\text{eq}} = (D_{\text{block}} \cdot C/n_{\text{total}})^{1/m}$ (from Basquin). This scalar S_{eq} allows direct comparison of different spectra and design for fatigue using constant-amplitude S-N data.

Deep Dive: Spectrum Loading and Statistical Extreme Values of Stress

Narrow-band Gaussian loading. If the stress process $\sigma(t)$ is a stationary, zero-mean Gaussian process with bandwidth parameter $\varepsilon \approx 0$ (narrow-band), the peaks are Rayleigh distributed: $f_P(s) = s/\sigma_\sigma^2 \exp(-s^2/(2\sigma_\sigma^2))$. The expected damage per peak (Basquin exponent m):

$$\mathbb{E}[d] = C^{-1}\mathbb{E}[S^m] = C^{-1}(2\sigma_\sigma^2)^{m/2}\Gamma(1 + m/2),$$

and the expected damage per unit time: $\mathbb{E}[\dot{D}] = \nu_0 \mathbb{E}[d]$, where ν_0 is the mean zero-crossing rate. This closed-form result is the *narrow-band approximation* for fatigue damage from a random Gaussian load.

Broad-band correction. For broad-band processes (ε not small), the narrow-band approximation overestimates damage because not all peaks form complete cycles. The Dirlik correction factor $\eta(\varepsilon) < 1$ corrects: $\mathbb{E}[\dot{D}]_{\text{BB}} = \eta(\varepsilon) \cdot \mathbb{E}[\dot{D}]_{\text{NB}}$. Dirlik's empirical formula (1985) is currently the most widely used broad-band fatigue damage estimate and is embedded in FEA software packages.

Extreme stress prediction. For a mission of duration T containing $n_{\text{peaks}} = \nu_0 T$ peaks, the expected maximum peak (annual maximum for $\nu_0 = 1$ Hz, $T = 31.5 \times 10^6$ s):

$$\mathbb{E}[\max S] \approx \sigma_\sigma \sqrt{2 \ln n_{\text{peaks}}} + \frac{\gamma_E \sigma_\sigma}{\sqrt{2 \ln n_{\text{peaks}}}},$$

where $\gamma_E = 0.5772$ is the Euler–Mascheroni constant. This follows from Gumbel extreme value theory for Rayleigh maxima (cf. Vol. I, Part 2). The annual maximum stress is the relevant loading for fracture assessment (Section 6.3).

Example — Annual maximum wind gust stress in a lattice tower

Wind process: stationary Gaussian, $\sigma_\sigma = 25$ MPa (RMS stress), $\nu_0 = 0.5$ Hz. Annual peaks: $n = 0.5 \times 3.15 \times 10^7 = 1.575 \times 10^7$.

$$\mathbb{E}[S^{1\,\text{yr}}_{\max}] \approx 25\sqrt{2\ln(1.575 \times 10^7)} + \frac{0.5772 \times 25}{\sqrt{2\ln(1.575 \times 10^7)}} = 25 \times 3.886 + \frac{14.4}{3.886} = 97.2 + 3.7 = 100.9\,\text{MPa}.$$

50-year return period stress: $S_{50} = \sigma_\sigma\sqrt{2\ln(n \times 50)} = 25\sqrt{2\ln(7.88 \times 10^8)} = 25 \times 4.310 = 107.7\,\text{MPa}$. Comparing to fatigue limit $S_e = 150\,\text{MPa}$: the annual maximum is well below the fatigue limit, but cumulative fatigue damage from the full distribution must still be checked via Miner's rule.

6.3 Paris' Law and Fatigue Crack Growth

Linear Elastic Fracture Mechanics and Paris' Law

Consider a crack of length a in a structure under cyclic loading with stress range $\Delta\sigma = \sigma_{\max} - \sigma_{\min}$. The *stress intensity factor range* is:

$$\Delta K = Y\Delta\sigma\sqrt{\pi a},$$

where Y is a dimensionless geometry factor (e.g. $Y = 1$ for a through-crack in an infinite plate). **Paris' Law** (Paris–Erdogan, 1963) states:

$$\frac{da}{dN} = C_P(\Delta K)^m,$$

where C_P and m are material constants ($m \approx 2$–4 for metals), and da/dN is the crack growth rate per cycle.

Three-region crack growth curve.

1. **Region I** ($\Delta K < \Delta K_{\text{th}}$): sub-threshold, negligible growth.
2. **Region II** ($\Delta K_{\text{th}} < \Delta K < K_c/\sqrt{Y^2\pi a_c}$): Paris regime, linear on log–log scale.
3. **Region III** ($\Delta K \to K_c = K_{Ic}$): unstable fracture. The NASGRO equation adds correction terms: $\frac{da}{dN} = C_P\frac{(1-f)^m\Delta K^m}{(1-\Delta K_{\text{th}}/\Delta K)^p(1-\Delta K_{\max}/K_c)^q}$.

Number of cycles to failure. Integrating Paris' law from initial crack a_0 to critical crack $a_c = (K_{Ic}/(Y\sigma_{\max}))^2/\pi$:

$$N_f = \int_{a_0}^{a_c}\frac{da}{C_P(Y\Delta\sigma\sqrt{\pi a})^m} = \frac{1}{C_P(Y\Delta\sigma)^m\pi^{m/2}}\int_{a_0}^{a_c}a^{-m/2}\,da.$$

For $m \neq 2$:

$$N_f = \frac{2}{(m-2)C_P(Y\Delta\sigma)^m\pi^{m/2}}\left[a_0^{1-m/2} - a_c^{1-m/2}\right].$$

For $m = 2$: $N_f = \frac{1}{C_P(Y\Delta\sigma)^2\pi}\ln(a_c/a_0)$.

Stress ratio and Walker equation. The effective ΔK depends on the stress ratio $R = \sigma_{\min}/\sigma_{\max}$. The Walker equation modifies Paris:

$$\frac{da}{dN} = C_P\left[\frac{\Delta K}{(1-R)^{1-\gamma}}\right]^m,$$

with $\gamma \in [0,1]$ a material parameter ($\gamma = 0$ recovers Paris; $\gamma = 1$ uses $K_{\max}$ instead of ΔK).

Example — Remaining life calculation for a pressure vessel

Data: $a_0 = 2$ mm (initial crack from NDT), $a_c = 25$ mm (critical crack for $K_{Ic} = 55$ MPa$\sqrt{\text{m}}$), $\Delta\sigma = 150$ MPa, $Y = 1.12$, $m = 3.5$, $C_P = 3 \times 10^{-12}$ (SI units, a in metres).
ΔK at a_0: $1.12 \times 150 \times \sqrt{\pi \times 0.002} = 168 \times 0.0792 = 13.3$ MPa$\sqrt{\text{m}}$.
Compute N_f with $m = 3.5$:

$$N_f = \frac{2}{(3.5-2) \times 3 \times 10^{-12} \times (1.12 \times 150)^{3.5} \pi^{1.75}} [0.002^{-0.75} - 0.025^{-0.75}].$$

$(1.12 \times 150)^{3.5} \approx 168^{3.5} \approx 1.65 \times 10^7$. $\pi^{1.75} \approx 7.07$. Denominator: $1.5 \times 3 \times 10^{-12} \times 1.65 \times 10^7 \times 7.07 \approx 2.09 \times 10^{-4}$.
Bracket: $0.002^{-0.75} \approx 102.5$; $0.025^{-0.75} \approx 17.4$. Bracket ≈ 85.1.
$N_f \approx 85.1/(2.09 \times 10^{-4}) \approx 4.1 \times 10^5$ cycles.

Stochastic crack growth. Real crack paths are random due to microstructural heterogeneity. The Paris parameters $\ln C_P$ and m are jointly normally distributed (negatively correlated). Monte Carlo propagation of 10^4 crack growth trajectories from a_0 gives a distribution of N_f, from which the B_{10} life (10th-percentile life) is extracted for design.

Stress intensity factor solutions. The geometry factor Y in $\Delta K = Y\Delta\sigma\sqrt{\pi a}$ depends on specimen geometry and loading:

- Through crack in infinite plate: $Y = 1$.
- Edge crack in semi-infinite plate: $Y \approx 1.12$.
- Embedded elliptical crack (semi-axes a, c): $Y = Q(a/c)^{-1/2}$ where $Q = \sqrt{1 - 1.464(a/c)^{1.65}}$ for $a < c$.
- Through-thickness crack in finite-width plate (width W): $Y = \sqrt{\sec(\pi a/W)}$ (Feddersen correction).

These solutions come from boundary integral equations or weight function methods.

Crack closure and the effective stress intensity. Under compressive loading or residual stress, crack faces can contact (close) before the minimum load, reducing the effective ΔK:

$$\Delta K_{\text{eff}} = K_{\max} - K_{\text{open}},$$

where K_{open} is the SIF at crack opening. Elber (1970) quantified plasticity-induced crack closure via an empirical factor $U = \Delta K_{\text{eff}}/\Delta K \approx 0.5 + 0.4R$ (for $0 \le R \le 0.7$). The modified Paris law $da/dN = C_P(\Delta K_{\text{eff}})^m$ better fits data across R-ratios.

LEFM validity. Paris' law applies in the regime of *linear elastic fracture mechanics* (LEFM), requiring the plastic zone $r_p = (1/6\pi)(K_{Ic}/\sigma_Y)^2$ (Irwin estimate) to be small compared to specimen dimensions: $r_p \ll (W - a)$ and $r_p \ll B$ (thickness). When plasticity is extensive (thin sheets or tough materials), the J-integral or CTOD ($\delta = K^2/(\sigma_Y E)$ for plane stress) replaces K.

J-Integral and δ-CTOD

The J-integral is a path-independent contour integral around the crack tip:

$$J = \oint_\Gamma \left(W_e n_1 - T_i \frac{\partial u_i}{\partial x_1} \right) ds,$$

where W_e is the strain energy density, T_i the traction vector, u_i displacements, n_1 the outward normal component in the crack propagation direction. For LEFM: $J = K^2/E'$ (plane stress: $E' = E$; plane strain: $E' = E/(1-\nu^2)$). The J-based Paris law $da/dN = C_J \Delta J^m$ extends crack growth to elastic-plastic conditions.

Probabilistic fracture mechanics (PFM). The key uncertain quantities are: initial crack size a_0 (log-normal from NDT POD curve), fracture toughness K_{Ic} (Weibull with $m \approx 4$–7 for ceramics; normal for metals), Basquin exponent scatter. Monte Carlo PFM:

1. Sample $(a_0^{(k)}, K_{Ic}^{(k)}, C_P^{(k)}, m^{(k)})$ from their joint distribution.
2. Integrate Paris' law to find $N_f^{(k)}$.
3. Estimate $P_f(N) = \#\{N_f^{(k)} \le N\}/N_{\text{MC}}$.

Example — PFM for an aircraft wing fitting

$a_0 \sim \text{LN}(\mu = 1.5\text{mm}, \sigma = 0.4\text{mm})$; $K_{Ic} \sim \mathcal{N}(55, 5^2)$ MPa$\sqrt{\text{m}}$; $\Delta\sigma = 120$ MPa; $Y = 1.12$; $m = 3.2$; $C_P = 5 \times 10^{-12}$.
10^5 Monte Carlo runs: $\bar{N}_f = 2.8 \times 10^5$ cycles, $\sigma_{N_f} = 9.4 \times 10^4$ cycles. Failure probability at $N = 1.5 \times 10^5$ cycles (design life): $\hat{P}_f = 4.3 \times 10^{-3}$.
At $N = 10^5$ cycles: $\hat{P}_f = 8 \times 10^{-5}$ — meets the 10^{-4} per flight hour damage tolerance requirement at 100-cycle flight increments.

Detection of Growing Cracks. The *Probability of Detection* (POD) function $P_{\text{det}}(a) = \Phi((\ln a - \mu_{\text{POD}})/\sigma_{\text{POD}})$ describes the probability that NDT detects a crack of size a. The $a_{90/95}$ (crack size detectable with 90% probability at 95% confidence) is the standard inspection capability metric. Inspection intervals are set so that $\Pr(\text{crack grows from } a_{90/95} \text{ to } a_c \text{ between inspections}) \le 10^{-5}$.

Elastic-plastic fracture mechanics (EPFM). When the plastic zone r_p is large relative to crack size, LEFM breaks down and EPFM is needed. Two key parameters:

J-integral: $J = -\partial U/\partial a$ (energy release rate for elastic-plastic material). For a Ramberg–Osgood material $\varepsilon/\varepsilon_0 = \sigma/\sigma_0 + \alpha(\sigma/\sigma_0)^n$, the J-dominated field near the crack tip is:

$$\sigma_{ij} = \sigma_0 \left(\frac{J}{\alpha \sigma_0 \varepsilon_0 I_n r} \right)^{1/(n+1)} \tilde{\sigma}_{ij}(\theta, n),$$

where I_n and $\tilde{\sigma}_{ij}$ are tabulated functions (HRR field). Fracture at $J = J_{Ic}$.

Crack tip opening displacement (CTOD): $\delta = \frac{J}{\sigma_0 m}$ for plane stress ($m \approx 1$) and $\delta = J/(2\sigma_0)$ for plane strain. The δ_c at initiation is a material property used in BS 7448 fracture mechanics assessments.

Fracture mechanics assessment diagram (FAD). The R6 (UK) and SINTAP FAD plots a non-dimensionalised load ratio $L_r = \sigma_{\text{ref}}/\sigma_Y$ against a toughness ratio $K_r = K/K_{\text{mat}}$ (or J/J_{mat}). The failure assessment line (FAL):

$$K_r^2 = \frac{L_r^2 + L_r^4}{2\mu} \cdot (1 + L_r^2)^{-1} + \frac{L_r^6}{2\mu(1 + L_r^2)},$$

separates the safe region (below FAL) from failure (above FAL). The FAD combines both brittle fracture (high K_r, low L_r) and plastic collapse (high L_r, low K_r) modes in a single diagram.

Example — FAD assessment of a pressure vessel weld flaw

A weld contains a surface crack: $a = 5$ mm, $2c = 30$ mm. Stress: $\sigma_{\text{membrane}} = 180$ MPa, $\sigma_{\text{bend}} = 40$ MPa. Material: $\sigma_Y = 350$ MPa, $\sigma_{UTS} = 480$ MPa, $K_{\text{mat}} = 110$ MPa$\sqrt{\text{m}}$.
Applied $K_I = F_m\sigma_m\sqrt{\pi a} + F_b\sigma_b\sqrt{\pi a} = 0.92 \times 180 \times \sqrt{\pi \times 0.005} + 0.80 \times 40 \times \sqrt{\pi \times 0.005}$
$= 0.92 \times 180 \times 0.1253 + 0.80 \times 40 \times 0.1253 = 20.7 + 4.0 = 24.7$ MPa$\sqrt{\text{m}}$.
$K_r = 24.7/110 = 0.225$. $L_r = \sigma_{\text{ref}}/\sigma_Y$; $\sigma_{\text{ref}} = 1.05 \times (\sigma_m + \sigma_b/2) = 1.05 \times 200 = 210$ MPa. $L_r = 210/350 = 0.60$.
FAD check at $(L_r, K_r) = (0.60, 0.225)$: FAL value at $L_r = 0.60$ is $K_r^{\text{FAL}} \approx 0.88$. Since $0.225 < 0.88$: **safe** (large margin). The flaw is acceptable; no repair needed.

Weld residual stresses. Welding introduces residual tensile stresses at the weld toe that significantly accelerate crack growth. The residual SIF contribution K_{res} adds algebraically to the applied K: $K_{\text{total}} = K_{\text{applied}} + K_{\text{res}}$. For a typical butt weld, $K_{\text{res}}(a) \approx 0.7\sigma_Y\sqrt{\pi a}$ (through-thickness residual profile). Post-weld heat treatment (PWHT) reduces residual stresses by factor of 3–5.

6.4 Stochastic Degradation: Gamma and Inverse Gaussian Processes

Process-Based Degradation Models

A *degradation process* $\{X(t), t \geq 0\}$ represents the cumulative damage or loss of performance in a component over time. Failure occurs at the first passage time $T = \inf\{t : X(t) \geq \ell\}$ for a critical threshold ℓ.
Gamma process $(\alpha(t), \beta)$: $X(t) - X(s) \sim \text{Gamma}(\alpha(t) - \alpha(s), \beta)$ for $t > s$, with independent increments. **Inverse Gaussian (IG) process**: $X(t) \sim \text{IG}(\mu(t), \lambda(t)^2/\mu(t))$ with monotone non-decreasing paths.

Gamma process properties. With mean function $\alpha(t)$ and rate β:

- $\mathbb{E}[X(t)] = \alpha(t)/\beta$, $\text{Var}[X(t)] = \alpha(t)/\beta^2$.
- Coefficient of variation: $\text{CV} = 1/\sqrt{\alpha(t)}$, decreasing in t (degradation becomes more predictable).
- Common choice: $\alpha(t) = ct^\eta$ (power-law mean, with $\eta > 1$ for accelerating degradation).
- Failure time CDF: $F_T(\ell) = \Pr(T \leq t) = \Pr(X(t) \geq \ell) = 1 - F_{\text{Gamma}}(\ell;\, \alpha(t), \beta)$, where F_{Gamma} is the Gamma CDF.

First-passage time of the IG process. The IG process has closed-form first-passage time distribution. If $X(t) = \mu t$ (linear drift) plus Brownian motion with variance $\sigma^2 t$ (i.e., an IG process with these parameters):

$$f_T(t) = \sqrt{\frac{\ell^2\lambda}{2\pi t^3}}\exp\left(-\frac{\lambda(t - \ell/\mu)^2}{2(\ell/\mu)^2 t}\right),$$

which is exactly the Inverse Gaussian distribution $T \sim \text{IG}(\ell/\mu,\, \lambda(\ell/\mu)^2)$. Thus: the IG process $\Rightarrow$ IG-distributed failure time, explaining the name.

MLE for the Gamma process. Given n components with degradation measurements $\{x_{ij} = X_i(t_j)\}_{j=1}^m$, the increments $\Delta x_{ij} = x_{ij} - x_{i,j-1}$ are independent $\text{Gamma}(\Delta\alpha_j, \beta)$ with $\Delta\alpha_j =$

$c(t_j^\eta - t_{j-1}^\eta)$. Log-likelihood:

$$\ell(c,\eta,\beta) = \sum_{i,j} \left[(\Delta\alpha_j - 1)\ln \Delta x_{ij} - \Delta x_{ij}\beta + \Delta\alpha_j \ln\beta - \ln\Gamma(\Delta\alpha_j) \right].$$

Gradient-based optimisation (e.g. L-BFGS) is standard; the score with respect to η involves the digamma function $\psi(\Delta\alpha_j)$.

Example — Gamma process for coating thickness degradation

A protective coating degrades by corrosion. Measurements on $n = 20$ steel panels at $t = 1, 2, 3, 5, 8$ years (mm remaining):
Fitted Gamma process: $\hat{c} = 0.04$, $\hat{\eta} = 1.2$ (slightly accelerating), $\hat{\beta} = 2.5$.
Mean degradation at $t = 8$ years: $\mathbb{E}[X(8)] = 0.04 \times 8^{1.2}/2.5 = 0.04 \times 10.56/2.5 = 0.169$ mm.
With initial coating thickness $\ell = 0.5$ mm (failure threshold), predicted median failure time: solving $F_{\text{Gamma}}(0.5; ct^\eta, \beta) = 0.5$ gives $t_{50} \approx 14.2$ years. The 5th-percentile life (design life at 95% reliability): $t_5 \approx 9.8$ years.

Competing degradation modes. When failure can be caused by any of K degradation processes $X_1(t), \ldots, X_K(t)$ with thresholds $\ell_1, \ldots, \ell_K$, the failure time is $T = \min_k T_k$. Under independence, $S(t) = \prod_k S_k(t)$. Dependent processes are modelled via copulas on the marginals F_{T_k}.

Wiener process degradation model — full development. The linear Wiener process $X(t) = x_0 + \mu t + \sigma W_t$ (drift $\mu > 0$, diffusion $\sigma > 0$) is the most tractable degradation model. Its properties:

- $X(t) \sim \mathcal{N}(x_0 + \mu t, \sigma^2 t)$ — symmetric, allowing decreases (not always physical).
- The failure time $T = \inf\{t : X(t) \geq \ell\}$ has IG distribution (Section 6.5).
- Degradation measurements at times $t_1 < \cdots < t_m$ yield increments $\Delta x_j = x(t_j) - x(t_{j-1}) \sim \mathcal{N}(\mu\Delta t_j, \sigma^2 \Delta t_j)$ — independent, enabling sequential MLE.

Random-effects Wiener model. Different units have different drift rates: $\mu_i \sim \mathcal{N}(\bar{\mu}, \tau^2)$. Conditional on μ_i: $\Delta x_{ij} \sim \mathcal{N}(\mu_i \Delta t_j, \sigma^2 \Delta t_j)$. The marginal increment distribution (integrating out μ_i) is $\mathcal{N}(\bar{\mu}\Delta t_j, \sigma^2\Delta t_j + \tau^2 \Delta t_j^2)$. The marginal failure time distribution no longer has a closed form but can be expressed as a scale mixture of IG distributions.

Updating predictions with observed degradation. Given m measurements $\{x_1, \ldots, x_m\}$ for unit i, the posterior for μ_i (with Gaussian prior $\mathcal{N}(\bar{\mu}, \tau^2)$) is:

$$\mu_i \,|\, \mathbf{x} \sim \mathcal{N}\left(\frac{\bar{\mu}/\tau^2 + (x_m - x_0)\sigma^{-2} t_m^{-1}}{1/\tau^2 + t_m/\sigma^2},\ \left(\frac{1}{\tau^2} + \frac{t_m}{\sigma^2} \right)^{-1} \right).$$

The updated RUL (remaining useful life) distribution is IG with updated drift $\hat{\mu}_i^{\text{post}}$:

$$\text{RUL}_i \sim \text{IG}\left(\frac{\ell - x_m}{\hat{\mu}_i^{\text{post}}},\ \frac{(\ell - x_m)^2}{\sigma^2} \right).$$

Example — Online RUL prediction for a Li-ion battery

Battery capacity $X(t)$ (fraction of nominal) degrades as a Wiener process with drift $\mu = -0.0003$ capacity/cycle, $\sigma = 0.002$, and random effects $\tau = 0.0001$. Failure threshold: $\ell = 0.8$ (80% capacity). After 500 cycles: $X(500) = 0.87$.
Prior $\bar{\mu} = -0.0003$. Posterior after 500 cycles: $\hat{\mu}_i = (-0.0003/\tau^2 + (0.87-1)/(-0.0003 \times 500) \times (-1)/\ldots)$. Simplified: the posterior updates toward the observed degradation rate ($-0.13/500 = -2.6 \times 10^{-4}$ capacity/cycle), slightly faster than the prior.
Remaining capacity to failure: $\ell - X(500) = 0.87 - 0.80 = 0.07$. Mean RUL: $0.07/|\hat{\mu}_i| \approx 0.07/2.65 \times 10^{-4} \approx 264$ additional cycles. 5th-percentile RUL: approximately 180 cycles.

Exponential dispersion models for degradation. The Gamma and IG processes are special cases of the broader class of *additive processes with independent stationary increments* (Lévy processes with non-negative increments). The Lévy–Khintchine representation:

$$\mathbb{E}[e^{-sX(t)}] = \exp\left(-t \int_0^\infty (1 - e^{-su})\nu(du)\right),$$

where ν is the Lévy measure. For the Gamma process: $\nu(du) = ce^{-\beta u}/u\, du$ (infinite activity, continuous paths). For the compound Poisson process: $\nu = \lambda F_J$ (finite activity, jump paths). This unification allows model selection based on observed path smoothness.

Multivariate degradation and copulas. When K degradation indicators are monitored simultaneously ($\mathbf{X}(t) = (X_1(t), \ldots, X_K(t))$), the joint model can be specified via a copula on the marginal IG or Gamma processes. The Clayton copula captures common-shock degradation (all channels accelerate together); the Gaussian copula captures gradual correlation buildup. System failure at $T = \min_k T_k$ where $T_k = \inf\{t : X_k(t) \geq \ell_k\}$.

Inspection-based maintenance decisions using degradation models. Under a Gamma process with threshold ℓ, the optimal replacement threshold d^* (replace if $X > d^*$ at inspection) and inspection interval Δ are jointly optimised:

$$\min_{d^*, \Delta} g(d^*, \Delta) = \frac{C_I + C_p P_p(d^*, \Delta) + C_f P_f(d^*, \Delta) + C_d D(d^*, \Delta)}{\tau(d^*, \Delta)},$$

where P_p, P_f are the probabilities of preventive and corrective replacement per cycle, D is the expected downtime cost, and τ is the expected cycle duration. This is a nonlinear 2D optimisation solved numerically.

Residual useful life (RUL) confidence intervals via propagation of model uncertainty. When degradation model parameters $\hat{\theta}$ are estimated from data, the parameter uncertainty contributes to RUL uncertainty:

$$\mathrm{Var}[\mathrm{RUL}] \approx \mathrm{Var}_{\mathrm{process}}[\mathrm{RUL} \mid \hat{\theta}] + \mathrm{Var}_{\hat{\theta}}[\mathbb{E}[\mathrm{RUL} \mid \theta]].$$

The first term is the irreducible process noise; the second is the epistemic uncertainty from finite data. Bayesian inference (computing the posterior $p(\theta \mid \text{data})$ and integrating) naturally accounts for both.

Example — Bayesian Gamma process for aircraft engine hot section

EHM (Engine Health Monitoring) data: 120 engines, each with 12 gas path measurements taken at 100-flight-cycle intervals. Degradation indicator: Exhaust Gas Temper-

ature (EGT) margin reduction (°C). Fitted Gamma process: prior $\bar{c} = 0.15$, $\bar{\eta} = 1.05$, $\bar{\beta} = 8$. Posterior (after first 6 measurements per engine): $\hat{c} = 0.18$, $\hat{\eta} = 1.12$, $\hat{\beta} = 7.4$. For a specific engine currently at $X(600) = 18\,°C$ margin reduction, with threshold $\ell = 50\,°C$: Mean RUL: $(50 - 18)/(\hat{c} \times 600^{\hat{\eta}}/\hat{\beta}) \approx 32/(0.18 \times 731/7.4) \approx 32/17.8 \approx 1.8$ (in units of 100-cycle measurement intervals) $= 180$ cycles. 5th-percentile RUL: 112 cycles (maintenance trigger at this threshold for reliability target).

Non-stationary degradation: change-point models. Degradation often accelerates at a change-point τ (e.g. onset of pitting corrosion, delamination initiation). The change-point model:

$$X(t) = \begin{cases} \mu_1 t + \sigma_1 W_t & t \le \tau, \\ X(\tau) + \mu_2 (t - \tau) + \sigma_2 (W_t - W_\tau) & t > \tau, \end{cases}$$

with $\mu_2 > \mu_1$ (acceleration post-change). The change-point τ is itself random ($\tau \sim F_\tau$, e.g. exponential), adding a discrete latent variable. Bayesian MCMC estimates $(\mu_1, \mu_2, \sigma_1, \sigma_2, F_\tau)$ jointly from observed degradation paths.

Non-stationary Gamma process with covariates. When the degradation rate depends on operational covariates $\mathbf{z}(t)$ (temperature, load, humidity), the mean function generalises to:

$$\alpha(t; \mathbf{z}) = c \int_0^t \exp(\boldsymbol{\gamma}^\top \mathbf{z}(s))\, ds,$$

a covariate-weighted cumulative exposure. This is the proportional degradation model: covariates accelerate or decelerate the Gamma process intensity. For constant covariates $\mathbf{z}$: $\alpha(t) = c \exp(\boldsymbol{\gamma}^\top \mathbf{z}) t^\eta$, which combines the power-law shape (index η) with an Arrhenius-type stress factor.

Mixture Gamma process. When the degradation population is heterogeneous (e.g. two quality grades of components), a mixture:

$$f_X(x; t) = \pi f_{\text{Gamma}}(x; \alpha_1(t), \beta_1) + (1 - \pi) f_{\text{Gamma}}(x; \alpha_2(t), \beta_2),$$

models latent subpopulations. The EM algorithm (E-step: posterior class probabilities per observation; M-step: update (α_k, β_k, π)) fits the model. Identification requires at least three measurement times per component.

Two-phase Gamma process. In practice, degradation often accelerates after a change-point. The two-phase model:

$$X(t) = \begin{cases} X_1(t) & 0 \le t \le \tau, \\ X_1(\tau) + X_2(t - \tau) & t > \tau, \end{cases}$$

where $X_1 \sim \text{Gamma}(c_1 t^{\eta_1}, \beta)$ and $X_2 \sim \text{Gamma}(c_2 t^{\eta_2}, \beta)$ with $c_2 > c_1$ (acceleration). The change-point τ may be: (a) fixed (known inspection trigger); (b) random ($\tau \sim \text{Exp}(\mu_\tau)$, latent variable estimated by EM); (c) estimated as a model parameter by profile likelihood over $\tau \in \{t_1, \dots, t_m\}$.

Bivariate degradation via correlated Gamma processes. For two simultaneously monitored degradation channels $(X_1(t), X_2(t))$:

- *Shared frailty*: $X_k(t) = Z \cdot \tilde{X}_k(t)$ where $Z \sim \text{Gamma}(1/\theta, 1/\theta)$ is a shared random effect. Marginals are Gamma; joint distribution is a bivariate Gamma with Kendall's $\tau_K = \theta/(2 + \theta)$.

- *Copula approach*: specify marginal Gamma processes and link them via a Clayton or Gaussian copula on the CDF values $(F_{X_1}(x_1;t), F_{X_2}(x_2;t))$.
- *Subordinated Lévy process*: $(X_1(t), X_2(t)) = (Y_1(L(t)), Y_2(L(t)))$ where $L(t)$ is a common Lévy subordinator (shared operational time) and Y_k are independent Gamma processes. Dependence parameter = variance of $L(t)$ relative to t.

System failure at $T = \min(T_1, T_2)$ where $T_k = \inf\{t : X_k(t) \geq \ell_k\}$.

Example — Bivariate Gamma process for aircraft engine health monitoring

Two EHM channels: $X_1(t)$ = exhaust gas temperature (EGT) margin decay (°C), $X_2(t)$ = oil debris count (particles/litre). Both modelled as Gamma processes; shared frailty parameter $\hat{\theta} = 0.42$ (Kendall's $\tau_K = 0.42/2.42 = 0.17$ — mild positive dependence). Marginal parameters: $\hat{c}_1 = 0.12$, $\hat{\eta}_1 = 1.08$, $\hat{\beta}_1 = 6$; $\hat{c}_2 = 0.08$, $\hat{\eta}_2 = 1.15$, $\hat{\beta}_2 = 4$. Failure thresholds: $\ell_1 = 50$°C, $\ell_2 = 200$ particles/litre. At $t = 800$ cycles: $\Pr(T_1 \leq 800) = 1 - F_\Gamma(50; 0.12 \times 800^{1.08}, 6) \approx 0.063$. $\Pr(T_2 \leq 800) = 1 - F_\Gamma(200; 0.08 \times 800^{1.15}, 4) \approx 0.041$.
Under independence: $\Pr(T_{\min} \leq 800) = 1 - 0.937 \times 0.959 = 0.101$. Under shared frailty: $\Pr(T_{\min} \leq 800) \approx 0.118$ (15% higher — dependence increases joint failure risk).

Degradation-based prognostics: the D-S-I-P framework.

1. *Data*: raw sensor measurements, cleaned and synchronised.
2. *Signature*: feature extraction — degradation indicator $X(t)$ constructed from raw data (e.g. RMS vibration, spectral kurtosis, oil viscosity).
3. *Inference*: fit the degradation process model to observed trajectory; update posterior parameters via Bayesian filtering.
4. *Prediction*: compute RUL distribution $p(\text{RUL}|\mathbf{x}_{1:m})$ and its percentiles for maintenance decision.

The D-S-I-P framework is implemented in ISO 13379-1 (condition monitoring data interpretation) and underpins commercial PHM (Prognostics and Health Management) systems.

6.5 First-Passage-Time Models for Failure

First Passage Time (FPT)

Given a stochastic process $\{X(t), t \geq 0\}$ and a threshold $\ell > 0$, the *first passage time* is:

$$T = \inf\{t \geq 0 : X(t) \geq \ell\}.$$

For diffusion processes $dX = \mu(X,t)\,dt + \sigma(X,t)\,dW_t$, the FPT density satisfies the *Fokker–Planck* and *Kolmogorov backward* equations.

Wiener process with drift. Let $X(t) = \mu t + \sigma W_t$ (Brownian motion with drift $\mu > 0$). The FPT to level $\ell > 0$ has the Inverse Gaussian distribution:

$$T \sim \text{IG}\left(\frac{\ell}{\mu}, \frac{\ell^2}{\sigma^2}\right), \quad f_T(t) = \frac{\ell}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{(\ell - \mu t)^2}{2\sigma^2 t}\right).$$

Proof sketch. By the reflection principle for Brownian motion: $\Pr(T \leq t) = \Phi\left(\frac{\mu t - \ell}{\sigma\sqrt{t}}\right) +$

$e^{2\mu\ell/\sigma^2}\,\Phi\big(\frac{-\mu t-\ell}{\sigma\sqrt{t}}\big)$. Differentiating with respect to t yields the IG density.

Ornstein–Uhlenbeck process. For $dX = -\theta X\,dt + \sigma\,dW_t$ (mean-reverting degradation), no closed-form FPT distribution exists. The FPT density $f(t; x_0, \ell)$ satisfies the backward Kolmogorov equation:

$$\frac{\partial f}{\partial t} = -\theta x \frac{\partial f}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial x^2},$$

with boundary condition $f = \delta(0)$ at $x = \ell$. Solved numerically by finite differences or via Laplace transforms (Siegert formula).

Random accelerated degradation. In practice, the degradation rate μ is itself random across units: $\mu_i \sim \mathcal{N}(\bar{\mu}, \tau^2)$. Conditional on μ_i, $T_i \sim \text{IG}(\ell/\mu_i, \ell^2/\sigma^2)$. The marginal failure time distribution is then a *normal–IG mixture*, with heavier tails than the pure IG.

Example — FPT for bearing vibration degradation

Vibration amplitude $X(t)$ (in g) of a rolling bearing follows a Wiener process with drift: $X(t) = 0.02t + 0.05W_t$, starting at $X(0) = 0.1\,g$. Failure threshold: $\ell = 1.0\,g$ (alarm level). Effective threshold for FPT from current state: $\ell^* = 1.0 - 0.1 = 0.9\,g$.
FPT distribution: $T \sim \text{IG}(\ell^*/\mu, (\ell^*)^2/\sigma^2) = \text{IG}(0.9/0.02, 0.9^2/0.0025) = \text{IG}(45, 324)$.
Mean time to failure: $\mathbb{E}[T] = 45$ time units. Variance: $\text{Var}[T] = (\ell^*)^3/(\mu^3 \cdot (\ell^*/\sigma)^2) = 324/(\mu^3) \to \text{Var}[T] = \mu/\lambda \cdot (\ell^*/\mu)^2 = 45^3/324 \approx 280$. Standard deviation: ≈ 16.7 time units.
Probability of failure before $t = 30$: $F_T(30) = \Phi\big(\frac{0.02\times 30-0.9}{0.05\sqrt{30}}\big) + e^{2\times 0.02\times 0.9/0.0025}\Phi\big(\frac{-0.02\times 30-0.9}{0.05\sqrt{30}}\big) \approx \Phi(-2.19) + e^{14.4}\Phi(-13.1) \approx 0.014$.

Two-boundary FPT. If failure can also occur by exceeding a lower threshold (e.g. voltage too low), we seek the FPT to either boundary $\ell^+ > X(0) > \ell^-$. For Brownian motion with drift, the FPT to $\{0, b\}$ starting at x has moment generating function:

$$\mathbb{E}_x[e^{-sT}] = \frac{\sinh(\sqrt{2s+\theta^2}(b-x)/\sigma)}{\sinh(\sqrt{2s+\theta^2}\,b/\sigma)} \cdot e^{\theta(b-x)/\sigma^2},$$

where $\theta = \mu/\sigma^2$ is the drift parameter. Inverted via numerical Laplace inversion (Euler method).

Part 7

Renewal Theory and Maintenance Models

Renewal processes underlying repairable systems, and the optimization of maintenance and replacement policies.

Brownian bridge and degradation with a target. If a component is replaced at a fixed inspection time T_{insp} regardless of the degradation level, the relevant process is the Brownian bridge $B_t^o = W_t - (t/T_{\text{insp}})W_{T_{\text{insp}}}$, conditioning the path to return to 0 at time T_{insp}. The maximum of a Brownian bridge:

$$\Pr\Big(\max_{0\le t\le T} B_t^o \le b\Big) = 1 - e^{-2b^2/T}, \quad b > 0.$$

This governs the probability of exceeding a threshold during the inspection interval.

Lévy processes and jump degradation. For degradation with occasional large jumps (e.g. shock-induced damage increments), the compound Poisson process $X(t) = \sum_{k=1}^{N(t)} J_k$ (with

$N(t) \sim \text{Poisson}(\lambda t)$ and i.i.d. jump sizes J_k) models each shock as a discrete damage increment. The FPT to ℓ:

$$T = \inf\left\{ t : \sum_{k=1}^{N(t)} J_k \geq \ell \right\},$$

is a compound distribution — a Poisson-stopped sum. If $J_k \sim \text{Exp}(\mu_J)$, the FPT has a closed-form Laplace transform; in general, it requires Monte Carlo or PGF inversion.

Hitting time asymptotics. For large threshold $\ell \to \infty$, the normalised FPT $T_\ell / \mathbb{E}[T_\ell]$ converges in distribution to a generalised inverse Gaussian distribution (for diffusions) or an exponential (for renewal-type processes). Wald's identity gives the mean: $\mathbb{E}[T_\ell] = \ell/\mu$ (for drift μ). The variance: $\text{Var}[T_\ell] = \ell\sigma^2/\mu^3$ (for Wiener), confirming that the coefficient of variation of the FPT $\sim (\ell\mu/\sigma^2)^{-1/2}$ decreases with threshold and drift.

Fokker–Planck Equation for FPT Density

For the diffusion $dX = \mu(X)\,dt + \sigma(X)\,dW_t$ starting at $x_0 < \ell$, the FPT density $p(t; x_0)$ satisfies the backward Kolmogorov (Feynman–Kac) equation:

$$\frac{\partial p}{\partial t} = \mu(x_0)\frac{\partial p}{\partial x_0} + \frac{\sigma^2(x_0)}{2}\frac{\partial^2 p}{\partial x_0^2},$$

with initial condition $p(0; x_0) = \delta(x_0 - \ell)$ (singular at the boundary) and $p(t; \ell) = 0$ for $t > 0$. In Laplace space $\tilde{p}(s; x_0) = \mathbb{E}[e^{-sT}]$:

$$\mu(x_0)\tilde{p}'(x_0) + \frac{\sigma^2(x_0)}{2}\tilde{p}''(x_0) = s\tilde{p}(x_0),$$

a second-order ODE with boundary conditions $\tilde{p}(\ell) = 1$ and $\tilde{p}(-\infty) = 0$ (or zero-flux at a lower boundary).

Numerical solution of the backward equation. For general drift/diffusion: discretise $x_0 \in [0, \ell]$ on a grid $\{x_j\}$, approximate derivatives by finite differences, and solve the resulting tridiagonal linear system for $\tilde{p}(s; x_j)$ at a given s. Invert the Laplace transform by the Euler numerical inversion algorithm (Abate–Whitt):

$$p(t; x_0) \approx \frac{e^{A/2}}{T}\left[\frac{1}{2}\tilde{p}\left(\frac{A}{2T}; x_0\right) + \sum_{k=1}^{N}(-1)^k \,\text{Re}\,\tilde{p}\left(\frac{A + 2\pi i k}{2T}; x_0\right)\right],$$

with $A = 18.4$, $N = 100$ giving errors $< 10^{-9}$ for smooth p.

Example — Ornstein–Uhlenbeck FPT for a corrosion model

Corrosion depth $X(t)$ follows an OU process: $dX = -\theta(X - \bar{X})\,dt + \sigma\,dW_t$ with $\theta = 0.05$ yr^{-1}, $\bar{X} = 3$ mm (mean corrosion depth), $\sigma = 0.5$ mm/yr$^{1/2}$, $x_0 = 1$ mm, $\ell = 6$ mm (perforation threshold).

The mean FPT (via the Siegert formula): $\mathbb{E}[T] = \int_{x_0}^{\ell} \frac{2}{\sigma^2(y)} e^{\Phi(y)} \int_{-\infty}^{y} e^{-\Phi(z)} dz\, dy$, where $\Phi(y) = \theta(y - \bar{X})^2/\sigma^2$.

Numerical evaluation: $\mathbb{E}[T] \approx 38$ years. The 5th-percentile FPT (from backward equation at $s = 0.03$): $t_5 \approx 21$ years. The 95th-percentile: $t_{95} \approx 67$ years — wide uncertainty due to the mean-reverting process exploring the region near $\bar{X} = 3$ mm for a long time before eventually crossing $\ell = 6$ mm.

Remaining useful life (RUL) distributions in practice. In condition monitoring, the degradation model is fitted sequentially: as new measurements arrive, the parameter posterior (and hence RUL distribution) is updated via Bayesian filtering (Kalman for Wiener; particle filter for non-linear/non-Gaussian models). The key challenge is the *tail of the RUL distribution* — the 1st or 5th percentile — because maintenance decisions based on mean RUL can be non-conservative. The 90% RUL confidence interval width typically halves with each doubling of observed degradation history.

Stochastic fatigue crack growth via SDE. The Paris equation under random loading becomes a stochastic differential equation:

$$da = C_P (Y\sigma_t\sqrt{\pi a})^m \, dN + \sigma_{\text{crack}}\sqrt{a}\, dW_N,$$

where dW_N is a Wiener increment in cycle count N and σ_{crack} models microstructural variability. The Fokker–Planck equation for the crack length PDF $p(a, N)$:

$$\frac{\partial p}{\partial N} = -\frac{\partial}{\partial a}[C_P(\Delta K)^m p] + \frac{\sigma_{\text{crack}}^2}{2}\frac{\partial^2 (ap)}{\partial a^2},$$

solved numerically by finite volumes. The variance of $a(N)$ grows with N: $\text{Var}[a(N)] \sim O(N^{2m/(m-2)})$ for the Paris exponent $m > 2$.

Importance sampling for rare FPT events. For high thresholds ℓ, $\Pr(T \le t)$ may be $< 10^{-6}$ — requiring $> 10^8$ crude Monte Carlo samples. Importance sampling uses a biased process $\tilde{X}(t)$ with larger drift ($\tilde{\mu} > \mu$) to generate more failure paths, then reweights:

$$\hat{P}_f = \frac{1}{N_{\text{IS}}}\sum_{k=1}^{N_{\text{IS}}} \mathbf{1}[T^{(k)} \le t] \cdot L^{(k)},$$

where $L^{(k)} = \exp\left[(\mu - \tilde{\mu})X(t^{(k)})/\sigma^2 - (\tilde{\mu}^2 - \mu^2)t^{(k)}/(2\sigma^2)\right]$ is the Radon–Nikodym derivative. Optimal $\tilde{\mu}$ minimises variance: $\tilde{\mu}^* = \mu + 2(\ell - x_0)/t$ (Doob's h-transform for Brownian motion).

Example — Importance sampling for structural reliability

A bridge girder: $X(t) =$ crack length (Wiener drift model), $\mu = 0.001$ mm/year, $\sigma = 0.01$ mm/yr$^{1/2}$, $x_0 = 0.5$ mm, $\ell = 5$ mm, target $P_f = \Pr(T \le 50\,\text{yr})$.
Crude MC: needs $\approx 10^6$ samples for $\hat{P}_f \approx 2 \times 10^{-4}$ with 30% coefficient of variation.
IS with $\tilde{\mu}^* = 0.001 + 2(4.5)/50 = 0.181$ mm/year: 10^3 IS samples give $\hat{P}_f = 2.3 \times 10^{-4}$ with 8% CV. A 125× variance reduction factor. The IS estimate is verified to be unbiased by checking $\mathbb{E}[L^{(k)}] = 1$ to within 0.5%.

FPT for multi-dimensional diffusions. If degradation is a vector $\mathbf{X}(t) \in \mathbb{R}^d$ (e.g. multiple corrosion pits on different structural elements), the failure domain is $\mathcal{F} = \{\mathbf{x} : g(\mathbf{x}) \le 0\}$ for a limit state function g. The multi-dimensional backward equation:

$$\boldsymbol{\mu}^\top \nabla_{\mathbf{x}_0}\tilde{p} + \frac{1}{2}\text{tr}[\Sigma \text{Hess}_{\mathbf{x}_0}\tilde{p}] = s\tilde{p},$$

where $\Sigma = \sigma\sigma^\top$ is the diffusion tensor. For correlated degradation processes (off-diagonal Σ), the FPT is shorter than for independent processes — correlations accelerate crossing of non-convex failure domains.

7.1 Renewal Processes and the Renewal Equation

Renewal Process

A *renewal process* $\{N(t), t \geq 0\}$ counts the number of failures (renewals) of a component in $[0, t]$, where successive inter-renewal times $X_1, X_2, \ldots$ are i.i.d. positive random variables with CDF F and mean $\mu = \mathbb{E}[X_i] < \infty$. The k-th renewal occurs at $S_k = \sum_{i=1}^{k} X_i$.

Renewal function. $M(t) = \mathbb{E}[N(t)] = \sum_{n=1}^{\infty} F^{*n}(t)$, where F^{*n} is the n-fold convolution of F. The renewal function satisfies the *renewal equation*:

$$M(t) = F(t) + \int_0^t M(t-s)\, dF(s).$$

Proof. Condition on the first renewal time X_1: $M(t) = \mathbb{E}[N(t)] = \mathbb{E}[N(t) \mid X_1 \leq t]\Pr(X_1 \leq t) + 0 = \int_0^t [1 + M(t-s)]\, dF(s) = F(t) + \int_0^t M(t-s)\, dF(s)$.

Elementary renewal theorem. As $t \to \infty$:

$$\frac{M(t)}{t} \to \frac{1}{\mu}.$$

Rate of renewals converges to $1/\mu$ regardless of the inter-renewal distribution. In practice, $M(t) \approx t/\mu$ for large t — used to estimate replacement frequency.

Key renewal theorem (Blackwell's theorem). For non-arithmetic F, for any $h > 0$:

$$M(t+h) - M(t) \to \frac{h}{\mu} \quad \text{as } t \to \infty.$$

For arithmetic distributions with span d: $M(nd) - M((n-1)d) \to d/\mu$.

Variance of $N(t)$. Using the renewal reward theorem:

$$\mathrm{Var}[N(t)] \approx \frac{\sigma^2 t}{\mu^3} \quad \text{for large } t,$$

where $\sigma^2 = \mathrm{Var}[X_i]$. The coefficient of variation of $N(t)$ decreases as $t^{-1/2}$.

Numerical solution. The renewal equation is solved numerically by discretisation: let $m_k = M(k\delta)$, $f_k = F(k\delta) - F((k-1)\delta)$. Then:

$$m_k = F(k\delta) + \sum_{j=1}^{k-1} m_{k-j} f_j \cdot \delta^{-1} \cdot \delta = \sum_{j=1}^{k} f_j + \sum_{j=1}^{k-1} m_{k-j} f_j.$$

This $O(k^2)$ recursion is exact in the limit $\delta \to 0$; for speed, the convolution $m = f * (f + m)$ is computed via FFT in $O(k \log k)$.

Example — Renewal function for a Weibull renewal process

Inter-renewal times $X_i \sim \text{Weibull}(\lambda = 0.1, k = 2)$ with $\mu = \Gamma(1+1/2)/0.1 = \sqrt{\pi}/0.2 \approx 8.86$ units.
Elementary renewal theorem: $M(t) \approx t/8.86$ for large t.
At $t = 100$: $M(100) \approx 100/8.86 \approx 11.3$ expected renewals.
Numerical solution at $t = 20$: $M(20) \approx 2.1$ (below the asymptotic rate of $20/8.86 = 2.26$, as is typical — initial sparseness of renewals).

Alternating renewal process. The system alternates between ON (duration $U_i \sim F_U$) and OFF (repair, $V_i \sim F_V$). Availability at time t:

$$A(t) = \Pr(\text{system ON at } t) \to \frac{\mu_U}{\mu_U + \mu_V} \quad \text{as } t \to \infty.$$

The asymptotic availability depends only on mean ON and OFF times — not on their distributions (a surprising consequence of the renewal reward theorem).

Spent renewal processes and delayed renewal. A *delayed renewal process* has the first inter-arrival time $X_1 \sim G \neq F$ (e.g. the "age" of the initial component is unknown). The renewal function satisfies $M_G(t) = G(t) + \int_0^t M(t-s)\, dG(s)$. For an *equilibrium renewal process*: $G(t) = F_e(t) = \mu^{-1} \int_0^t S(u)\, du$ (the equilibrium distribution). The equilibrium renewal process is stationary: $M_G(t) = t/\mu$ exactly (not just asymptotically).

Inspection paradox. In a renewal process, the length of the current interval (at a random inspection time t) has distribution stochastically larger than F:

$$\Pr(X_{\text{current}} > x) = \frac{\int_x^\infty S(u)\, du}{\mu} \cdot 2 - \frac{S(x)\mu^{(2)}}{2\mu^2\,?} \ldots$$

More precisely, the length $L(t) = S_{N(t)+1} - S_{N(t)}$ satisfies $\mathbb{E}[L(t)] \to \mu + \sigma^2/\mu$ as $t \to \infty$ (always $\geq \mu$, with equality only if $\sigma = 0$). This is the *inspection paradox*: buses arrive at rate $1/\mu$, yet a random passenger waits $\approx \mu/2 + \sigma^2/(2\mu)$ on average — longer for high-variance inter-arrival times.

Renewal theory and queueing. The M/G/1 queue's mean busy period $\mathbb{E}[B] = \rho/(1-\rho)$ (where $\rho = \lambda/\mu$ is the traffic intensity) follows directly from renewal reward theory applied to busy-idle cycles. The Pollaczek–Khinchine formula for mean queue length is a classic renewal reward result.

Central Limit Theorem for Renewal Processes

As $t \to \infty$:

$$\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \xrightarrow{d} \mathcal{N}(0,1),$$

where $\sigma^2 = \text{Var}[X_i]$. This gives a normal approximation for the number of renewals and enables confidence intervals on predicted maintenance counts over long horizons.

Moment approximations for $M(t)$. The Smoluchowski approximation for small t:

$$M(t) \approx \frac{F(t)}{1 - F(t)/2}, \quad t \text{ small.}$$

The asymptotic expansion:

$$M(t) = \frac{t}{\mu} + \frac{\sigma^2 - \mu^2}{2\mu^2} + \frac{\mu_3'}{6\mu^3} \cdot \frac{1}{t} + O(t^{-2}),$$

where μ_3' is the third moment of F. For the Weibull with $k = 2$: $M(t) = t/\mu + (\sigma^2 - \mu^2)/(2\mu^2) \approx t/\mu - 0.215$ (the constant term is negative, reflecting that Weibull-2 is IFR — renewals occur less often than the asymptote early on).

Example — Hospital equipment replacement planning

Infusion pump MTTF: $\mu = 18$ months (Weibull, $k = 2.2$, $\sigma = 8$ months). Fleet size: $N = 50$ pumps. Expected replacements in 5 years (60 months): $50 \times M(60)$. Asymptotic: $M(60) \approx 60/18 - (64 - 324)/(2 \times 324) \approx 3.33 + 0.40 = 3.73$. Total: ≈ 187 replacements. Budget reserve (2-sigma): $50\sqrt{60 \times 64/18^3} = 50\sqrt{0.74} \approx 43$ additional units (a 23% safety margin).

Superimposed renewal processes. n independent renewal processes (components running in parallel) superimpose to give $N_{\text{total}}(t) = \sum_{i=1}^{n} N_i(t)$. $\mathbb{E}[N_{\text{total}}(t)] = nM(t)$. For large n, by the Palm–Khintchine theorem, the superimposed process converges to a Poisson process with rate n/μ — the basis of NHPP (non-homogeneous Poisson process) failure models for large fleets.

Superposition and thinning of renewal processes. The superposition of n independent renewal processes with i.i.d. inter-arrival distributions F converges, after rescaling, to a Poisson process. The *thinning* operation: independently remove each event with probability $1 - p$; the resulting process has inter-renewal distribution G satisfying $\tilde{G}(s) = p\tilde{F}(s)/(1 - (1-p)\tilde{F}(s))$ (geometric thinning). These operations are fundamental to the NHPP approximation for large fleets.

Renewal equations: explicit solutions. For special F:

- *Exponential* $F(t) = 1 - e^{-\lambda t}$: $M(t) = \lambda t$ (exact, not just asymptotic).
- *Erlang-k* (k exponential stages in series): $M(t) = \lambda t/k + (k-1)/(2k) + \sum_{j=1}^{k-1}(\ldots)e^{-j\lambda t/k}$ — oscillates around the asymptote with period k/λ.
- *Gamma*(α, β): $M(t)$ is given by the Laplace inversion of $\tilde{F}(s)/(1 - \tilde{F}(s))$ numerically; the asymptotic correction term is $(\alpha^2 + \alpha - 1)/(2\alpha^2)$ (from $\mathbb{E}[X^2]$).

Renewal processes in discrete time. A discrete renewal process has inter-arrival distribution $\{p_k\}_{k=1}^{\infty}$ on $\mathbb{N}_+$. The renewal mass function $u_n = \Pr(S_k = n \text{ for some } k)$ satisfies:

$$u_n = p_n + \sum_{j=1}^{n-1} u_j p_{n-j}, \quad n \geq 1; \quad u_0 = 1.$$

The Markov chain with transition matrix $P_{ij} = p_{j-i+1}$ (for $j \geq i$) embeds the renewal process and allows matrix-analytic computation of transient probabilities.

Example — Discrete renewal for a weekly maintenance schedule

A machine is inspected weekly; it fails in week k with probability $p_k = (0.3)(0.7)^{k-1}$ (geometric, $\mu = 1/0.3 \approx 3.33$ weeks). Expected replacements in 10 weeks: $M(10) = 10\mu^{-1} + \ldots = 10/3.33 = 3.0$ (asymptotic). Exact via convolution: $M(10) = 2.96$ (close to asymptote; geometric has low variance).
Compare Poisson approximation ($n \to \infty$ superposition of 20 machines): expected failures per week $= 20/3.33 = 6.0$, variance ≈ 6.0 (Poisson). 95% prediction interval for weekly failures: $6.0 \pm 1.96\sqrt{6} = [1.2, 10.8]$.

Renewal process with rewards: applications. The renewal reward theorem enables closed-form long-run averages for complex systems:

- *Long-run fraction of time in state A*: $\lim_{t\to\infty} \Pr(\text{in state } A) = \mathbb{E}[\text{time in } A \text{ per cycle}]/\mu$.

- *Long-run rate of event occurrences*: $\lim_{t\to\infty} E[\text{events}]/t = \mathbb{E}[\text{events per cycle}]/\mu$.
- *Long-run mean reward rate*: $\lim_{t\to\infty} \mathbb{E}[\int_0^t r(X_s)\,ds]/t = \mathbb{E}[\int_0^\mu r(X_s)\,ds]/\mu$ (where X_s is the process state).

Regenerative processes. A process $\{X(t)\}$ is *regenerative* if there exist random times $0 = T_0 < T_1 < T_2 < \ldots$ (renewal epochs) at which the process probabilistically restarts. All functionals of X satisfy the renewal reward theorem. Under mild conditions (non-arithmetic F): $\pi(A) = \lim_{t\to\infty} \Pr(X(t) \in A) = \mathbb{E}[\int_0^\mu 1_{X_s\in A}\,ds]/\mu$. This is the basis for computing steady-state availability, utilisation rates, and long-run quality metrics.

7.2 Renewal-Reward Processes and Long-Run Costs

Renewal Reward Theorem

In a renewal reward process, let R_k be the reward earned in the k-th renewal cycle (of duration X_k). If $\mathbb{E}[|R_k|] < \infty$ and $\mathbb{E}[X_k] < \infty$, then:

$$\frac{\text{total reward in } [0,t]}{t} \to \frac{\mathbb{E}[R_k]}{\mathbb{E}[X_k]} \quad \text{w.p.1 and in } L^1 \text{ as } t\to\infty.$$

Application to maintenance cost. Consider a component replaced at failure (cost C_f) or preventively at age T_p (cost $C_p < C_f$). Each cycle either ends in failure ($X_k < T_p$, cost C_f) or planned replacement ($X_k = T_p$, cost C_p). Expected cycle cost:

$$\mathbb{E}[R] = C_f \Pr(T < T_p) + C_p \Pr(T \ge T_p) = C_f F(T_p) + C_p[1 - F(T_p)].$$

Expected cycle length:

$$\mathbb{E}[X] = \int_0^{T_p} t\,dF(t) + T_p[1-F(T_p)] = \int_0^{T_p} S(t)\,dt.$$

Long-run cost rate: $g(T_p) = \mathbb{E}[R]/\mathbb{E}[X]$.

Optimisation. Setting $dg/dT_p = 0$ gives the implicit equation for the optimal preventive replacement age T_p^*:

$$h(T_p^*)\cdot\int_0^{T_p^*} S(t)\,dt - F(T_p^*) = \frac{C_p}{C_f - C_p}.$$

This has a unique finite solution when $h(t)$ is IFR. If $h(t)$ is constant (exponential), $T_p^* = \infty$ — preventive replacement never optimal.

Discounted cost. With continuous discounting at rate $r > 0$, the present value of total cost starting with a new component is:

$$V = \mathbb{E}\left[\int_0^\infty e^{-rt} c\,dN(t)\right],$$

where c is the unit replacement cost. The present-value renewal equation: $V = \int_0^\infty e^{-rs}[c + V_{\text{new}}]dF(s)$, giving $V = c\,\tilde{F}(r)/(1-\tilde{F}(r))$ where $\tilde{F}(r) = \int_0^\infty e^{-rs}dF(s)$ is the Laplace-Stieltjes transform of F.

Example — Optimal block replacement policy

A machine fleet uses a component with Weibull failure time ($k = 2.5$, $\mu = 1000$ h). Failure replacement cost $C_f = \$500$; planned replacement cost $C_p = \$80$. Find T_p^*.
$F(t) = 1 - e^{-(\mu^{-1}\Gamma(1+1/k)t)^k}$. With $\Gamma(1.4) \approx 0.887$: $\lambda = 0.887/1000 = 8.87 \times 10^{-4}$.
$h(t) = k\lambda^k t^{k-1} = 2.5 \times (8.87 \times 10^{-4})^{2.5} t^{1.5}$.
Numerically solving $g'(T_p) = 0$: $T_p^* \approx 720$ h.
At T_p^*: $F(720) \approx 0.21$, $g(720^*) \approx (0.21 \times 500 + 0.79 \times 80)/\int_0^{720} S(t)dt \approx 183/627 \approx \$0.292$/h.
Comparison: failure-only policy (no PM): $g(\infty) = C_f/\mu = 500/1000 = \$0.50$/h. Savings: 41.6%.

Opportunistic maintenance. In multi-component systems, one component's planned downtime provides an opportunity to replace other components at reduced setup cost C_{setup}. The optimal opportunistic replacement policy balances individual T_p^* against the savings from coordinated downtime.

Preventive replacement under discounting. When future costs are discounted at rate r, the long-run cost criterion is replaced by the expected net present value (NPV). For an age replacement policy, the NPV equation:

$$V(x) = \int_0^{T_p - x} e^{-rs}\big[c_r + V(0)\big] f(s+x)/S(x)\, ds + e^{-r(T_p - x)}\big[c_p + V(0)\big] S(T_p)/S(x),$$

where $V(x)$ is the expected NPV for a component currently of age x, c_r is failure replacement cost, c_p is preventive replacement cost. Solving: $V(0)$ satisfies a fixed-point equation. As $r \to 0$, $V(0) \to g(T_p)/r$ (consistent with the long-run cost rate).

Multi-unit replacement: Group replacement. For n identical independent units, group replacement at age T_p reduces setup cost amortised over n units but causes premature replacement of units with remaining life. The cost savings from group replacement vs. individual are analysed via the theory of order statistics: the expected age of the youngest unit in a group of n at time T_p is $T_p - \mathbb{E}[T_{(1)}]$ where $T_{(1)}$ is the minimum of n i.i.d. lifetimes.

Repair/replace decision under increasing repair cost. Let the cost of the k-th repair be $c_k = c_1 q^{k-1}$ (geometric escalation, $q > 1$). Under the geometric process (Section 7.4), the optimal number of repairs before replacement N^* minimises the long-run cost:

$$g(N) = \frac{c_R + \sum_{k=1}^{N} c_k}{\sum_{k=1}^{N} \mathbb{E}[X_k]} = \frac{c_R + c_1(q^N - 1)/(q-1)}{\mu \sum_{k=1}^{N} a^{-(k-1)}},$$

where c_R is the replacement cost. Setting $g(N^*) \le g(N^* \pm 1)$ gives the optimal integer N^*.

Stochastic Control Formulation of Optimal Maintenance

Define the state (t, x) = (system age, observable condition). The optimal policy π^* minimises the total discounted expected cost:

$$J^{\pi}(t,x) = \mathbb{E}^{\pi}\left[\int_0^{\infty} e^{-rs} c(\text{state}, \text{action})\, ds \,\middle|\, \text{state} = (t,x)\right].$$

The Hamilton–Jacobi–Bellman (HJB) equation:

$$rJ^* = \min\Big\{ c_{\text{do nothing}} + \mathcal{A}J^*, \quad c_{\text{replace}} + J^*(0,0) - J^*(t,x) \Big\},$$

where $\mathcal{A}$ is the infinitesimal generator of the state process. For a Wiener degradation: $\mathcal{A}J = \mu\partial J/\partial x + (\sigma^2/2)\partial^2 J/\partial x^2$. The optimal policy has a *control limit* structure: replace when degradation exceeds a threshold $x^*(t)$ that solves the HJB boundary condition.

Example — Optimal maintenance for a pump with degradation measurements

State: residual capacity $x \in [0, x_0]$, declining Wiener process with $\mu = -0.01$ unit/day, $\sigma = 0.05$. Failure at $x = 0$. Replacement at cost $c_R = \$5000$; failure penalty $c_F = \$20000$; preventive replacement at any $x > 0$ costs c_R.
HJB solution (numerical, 2D grid on (t, x)): the optimal policy is a decreasing boundary $x^*(t)$ — as the component ages, the threshold for premature replacement increases (wait longer before replacing a young component). At $t = 0$: replace if $x < 0.15x_0$. At $t = 100$ days: replace if $x < 0.42x_0$ (threshold rises with age). The optimal policy saves 22% cost vs. the best fixed-threshold policy.

Monte Carlo simulation of maintenance cost. For complex systems where analytical solutions are intractable, Monte Carlo maintenance cost simulation is the practical tool: simulate the failure/repair history of each component over a planning horizon, applying the maintenance policy, and aggregate costs. Variance reduction techniques (control variates, antithetic sampling) reduce simulation error. Sensitivity analysis via automatic differentiation through the simulation computes $\partial\mathbb{E}[\text{cost}]/\partial\theta$ for policy parameter optimisation.

Monotone maintenance policies. A policy π is *monotone* if the replacement decision is non-decreasing in the component age (or degradation level). Under IFR failure distributions, the optimal age replacement policy $\pi^*(t) = \mathbf{1}[t \geq T_p^*]$ is monotone. This motivates the practical *threshold policies* where a decision rule is fully characterised by a single number.

Optimal inspection and replacement under partial observability (POMDP). When the component state is not directly observed (only inspections reveal information), the problem is a Partially Observable Markov Decision Process (POMDP). The state is the *belief* $b(s) = \Pr(\text{true state} = s)$, updated by Bayes' theorem after each inspection outcome. The optimal policy maps beliefs to actions (inspect, maintain, replace). For continuous degradation: the belief state is a probability distribution over degradation levels, updated via the Kalman or particle filter.

Optimal stopping and free boundary problems. The age replacement problem is an *optimal stopping problem*: choose a stopping time τ (the replacement time) to minimise $\mathbb{E}[c(\tau)]$. For the discounted criterion:

$$V(x) = \inf_{\tau} \mathbb{E}_x\left[\int_0^\tau e^{-rs} c_{\text{operating}}\, ds + e^{-r\tau} c_{\text{replace}}\right],$$

the optimal policy has a *free boundary*: replace when $x \geq x^*$, where x^* solves the smooth-pasting condition $V(x^*) = c_{\text{replace}}/r$ and $V'(x^*) = 0$ (smooth paste). This is a Stefan-type free boundary problem in the theory of stochastic control.

Example — Discounted age replacement for a generator

Failure time $T \sim \text{Weibull}(k = 2, \mu = 1000\,\text{h})$. Failure cost $c_f = \$10000$; replacement cost $c_r = \$2000$; operating cost rate $c_{\text{op}} = \$1/\text{h}$. Discount rate $r = 0.05/\text{year} = 5.7 \times 10^{-5}/\text{h}$.
Discounted cost function: $V(T_p) = \frac{c_r + (c_f - c_r)F(T_p) + c_{\text{op}} \int_0^{T_p} S(t)\,dt}{1 - e^{-r\mu_{\text{cycle}}(T_p)}}$ (approximate, for small

r).
Numerical optimisation: $T_p^* = 820\,\text{h}$, $V(T_p^*) = \$112000$ (NPV over 5 years). Compare with no PM: NPV $= \$145000$ (29% higher cost).

Network of maintenance policies. For a system with K components, the joint replacement policy $\pi(t_1, \ldots, t_K) \to \{\text{replace subset} \mathcal{S}\}$ is exponentially complex. *Opportunistic maintenance* exploits the system's planned downtime: when component i is replaced, inspect and replace components j with age $a_j \geq a_j^{\text{opp}}$, saving setup cost. The optimal opportunistic threshold a_j^{opp} is derived from the marginal cost equation:

$$g_j(a_j^{\text{opp}}) = c_{\text{opp,setup}} + c_j \cdot \mathbb{E}[\text{additional life gained by replacing at } a_j^{\text{opp}}],$$

where $c_{\text{opp,setup}} < c_{\text{standalone,setup}}$ is the reduced setup cost when piggybacking on another component's replacement.

Long-run optimal replacement with inflation. When costs increase at inflation rate π per unit time, the cost of the k-th replacement cycle (starting at calendar time S_k) is $C_f e^{\pi S_k}$ (at failure) or $C_p e^{\pi S_k}$ (at planned replacement). The optimal policy under discounting r and inflation π replaces the effective discount rate: $r_{\text{eff}} = r - \pi$. For $\pi < r$ (inflation below discount rate), the optimal T_p^* shifts upward — delay replacement because future costs are cheaper in real terms.

Capital replacement under technological improvement. If component quality improves over time (a newer replacement has longer life expectancy), the optimal replacement time accounts for the opportunity cost of waiting for better technology. The Terborgh economic life model balances increasing maintenance costs of the old component against the capital cost advantage of the new (better) technology.

Non-stationary reward rates. If the reward rate (revenue generated per unit time) depends on system age $r(t)$ (e.g. machine accuracy decreases with age), the renewal-reward theorem gives long-run profit rate:

$$\rho(T_p) = \frac{\int_0^{T_p} r(t) S(t)\,dt - C_p[1 - F(T_p)] - C_f F(T_p)}{\int_0^{T_p} S(t)\,dt}.$$

The optimal T_p^* maximises ρ, not necessarily identical to the cost-minimising T_p from Section 7.2.

Example — Profit-maximising replacement for a CNC machining centre

A CNC machine generates revenue at rate $r(t) = r_0(1 - \alpha t)$ (linear decrease due to precision loss). $r_0 = \$200$/h, $\alpha = 0.0002$/h. Failure distribution: Weibull($k = 2$, $\mu = 5000$h). Replacement costs: $C_f = \$15{,}000$, $C_p = \$5{,}000$.
Without revenue degradation: $T_p^* = 3840$ h, profit rate $\rho = \$190$/h. With $r(t) = r_0(1 - 0.0002t)$: the optimal replacement advances to $T_p^* = 3120$ h (replace sooner to avoid precision loss) and $\rho = \$184$/h. The 18% earlier replacement is economically justified by the $\$8$/h revenue loss at age 3840h vs. 3120h.

Ergodic theory and long-run averages. The renewal-reward theorem is a special case of the *ergodic theorem* for regenerative processes. Let $\{X(t)\}$ be any measurable functional of a regenerative process with cycles of mean length μ. Then:

$$\frac{1}{T}\int_0^T g(X(t))\,dt \to \frac{\mathbb{E}[\int_0^\tau g(X(t))\,dt]}{\mu} \quad \text{a.s.}$$

This underpins the validity of replacing time averages by ensemble averages in all reliability cost calculations, not just simple renewal processes.

Multi-objective replacement. In practice, replacement decisions involve multiple objectives: cost, availability, safety, and environmental impact. Pareto-optimal policies lie on the efficient frontier in objective space. The ε-constraint method converts all but one objective into constraints: minimise cost subject to availability $\geq a^*$ and safety failure rate $\leq \lambda^*$. The Pareto frontier is swept by varying (a^*, λ^*).

Game-theoretic maintenance. When multiple stakeholders share infrastructure (e.g. several airlines using the same airport taxiway), each has incentive to delay maintenance (free-riding on others' replacements). The Nash equilibrium replacement strategy satisfies: no player can reduce their cost by unilaterally changing their replacement time. Under certain conditions, the Nash equilibrium leads to under-maintenance relative to the social optimum — motivating centralised maintenance contracts or regulatory mandates.

7.3 Age and Block Replacement Policies

Age vs Block Replacement

Age replacement: replace at failure OR when component age reaches T_p (whichever comes first). Each cycle restarts from age 0.
Block replacement: replace all components at fixed times kT_p ($k = 1, 2, \ldots$) regardless of age, plus minimal repair at intermediate failures.

Age replacement cost rate. As derived in Section 7.2:

$$g_{\text{age}}(T_p) = \frac{C_p + (C_f - C_p)F(T_p)}{\int_0^{T_p} S(t)\,dt}.$$

Block replacement cost rate. At each block of length T_p, there are $M(T_p)$ failures (repaired minimally at cost c_f each) plus one planned replacement at cost C_p:

$$g_{\text{block}}(T_p) = \frac{C_p + c_f M(T_p)}{T_p}.$$

Note c_f (minimal repair cost) $\neq C_f$ (replacement cost): minimal repair restores the component to working order without rejuvenating it.

Comparison. Age replacement is generally more cost-efficient than block replacement because it avoids replacing relatively young components. However, block replacement is operationally simpler (scheduled maintenance windows, e.g. aircraft C-checks at fixed flight hours). The gap narrows for highly variable failure times (large σ/μ).

(n, T) **policy.** A generalisation: replace at failure or at age T (whichever comes first), but only if the component has already been replaced $n - 1$ times (accounting for fleet heterogeneity). Optimised over (n, T) jointly.

Example — Block vs age replacement for an exponential distribution

$X \sim \text{Exp}(\lambda = 0.001)$, $\mu = 1000$ h, $C_p = \$100$, $c_f = \$400$ (minimal repair), $C_f = \$400$ (replacement).
Age policy: $g'_{\text{age}}(T_p) = 0$ with exponential F (constant h) — since $h(t) = \lambda$ is IFR

only at the boundary case, the optimal age is $T_p^* = \infty$ (no PM beneficial). $g_{\text{age}}(\infty) = C_f/\mu = 400/1000 = \$0.40/\text{h}$.
Block policy: $M(T_p) \approx \lambda T_p = T_p/1000$. $g_{\text{block}}(T_p) = (100 + 400 \cdot T_p/1000)/T_p = 100/T_p + 0.40$. This is minimised at $T_p \to \infty$: $g_{\text{block}} \to \$0.40/\text{h}$.
Both policies reduce to failure-only replacement at the same rate, confirming that preventive replacement is useless for memoryless components.

Condition-based maintenance (CBM). Rather than age- or time-based triggers, CBM triggers replacement when a monitored degradation indicator $X(t)$ crosses a threshold $d^* < \ell$ (pre-alarm level). CBM strictly dominates age/block policies in expected cost for stochastically increasing degradation, because it replaces only when degradation evidence justifies it. The optimal threshold d^* balances the cost of premature replacement against the risk of failure before scheduled inspection.

Periodic inspection with imperfect detection. If failures are latent (not immediately revealed) and inspections detect them with probability $p_{\text{det}} < 1$, the optimal inspection interval τ^* balances the cost of missed failures against inspection cost:

$$g(\tau) = \frac{c_I + c_F P_f(\tau)(1 - p_{\text{det}}) \cdot \mathbb{E}[T_{\text{undetected}} \mid \text{undetected}]}{T},$$

where $P_f(\tau)$ is the probability of failure before inspection at τ and the expected "downtime" while failed-but-undetected grows with τ. Under perfect detection ($p_{\text{det}} = 1$): standard Weibull/exponential age-replacement. Under $p_{\text{det}} < 1$: optimal τ^* decreases (more frequent inspections needed).

Condition-based vs. time-based: CBM superiority. Under a Gamma degradation process $X(t)$ with threshold ℓ, the optimal CBM policy replaces when $X(t) \geq d^* < \ell$. The CBM cost rate is always $\leq$ the optimal age-replacement cost rate (Grall et al., 2002), with the gap increasing with degradation variance. Intuitively, CBM "knows" which units are close to failure and replaces them early (cheaply), while ignoring healthy units.

Predictive maintenance and the digital twin. A *digital twin* is a real-time Bayesian state estimator for the degradation process, running in parallel with the physical asset. At each sensor reading $y_k = X(t_k) + \varepsilon_k$, the posterior $p(X(t_k) \mid y_1, \ldots, y_k)$ is updated via Kalman filtering (linear Gaussian) or particle filtering (general). The RUL distribution $p(\text{RUL} \mid y_1, \ldots, y_k)$ is propagated forward. Maintenance is triggered when $P(\text{RUL} < \tau_{\text{threshold}}) > \alpha$.

Example — Block vs. age vs. CBM: lifecycle cost comparison

Fleet of 100 gearboxes. Failure distribution: Weibull($k = 2.5$, $\mu = 2000$ h). Failure cost: $c_f = \$8000$; planned replacement: $c_p = \$1200$; sensor installation: $\$200$/unit (one-time). Degradation model: Gamma process, $\sigma^2/\mu_{\text{deg}} = 0.4$ (moderate scatter).

Policy	Optimal parameter	Cost/h/unit
Failure-only	—	$\$4.00$
Block replacement	$T^* = 1350$ h	$\$2.68$
Age replacement	$T_p^* = 1180$ h	$\$2.39$
CBM	$d^* = 0.72\ell$	$\$2.01$

CBM reduces cost by 50% vs failure-only, and by 16% vs optimal age replacement. The $\$200$ sensor investment pays back in < 200 operating hours.

Opportunistic and group maintenance. When n components are in the same machine (e.g. bearings in a motor), it may be cost-effective to replace all when one is replaced (group

replacement), despite different individual ages. The *opportunistic maintenance* threshold: replace component i at age a_i opportunistically if $a_i > a^*_{\text{opp}}$, where a^*_{opp} satisfies a marginal cost equation comparing the cost of immediate replacement vs. waiting until the component's own optimal age.

Spare parts inventory integration. Optimal maintenance intervals interact with spare parts stock levels. The (s, Q) inventory policy for spare parts (order Q units when stock $\leq s$) must be jointly optimised with the replacement interval to avoid spare parts shortages (which extend downtime) or excess inventory (holding cost). The joint cost function:

$$g(\tau, s, Q) = g_{\text{maintenance}}(\tau) + h\mathbb{E}[\text{inventory}](s, Q, \lambda(\tau)) + c_{\text{order}}/Q \cdot \lambda(\tau),$$

where $\lambda(\tau) = 1/(\tau + \mu_{\text{repair}})$ is the demand rate for spares.

Periodic preventive maintenance with minimal repair at failure. A system is minimally repaired at each failure within a block $[kT_p, (k+1)T_p)$ and preventively restored (AGAN) at times kT_p. Under an NHPP failure process with intensity $\rho(t)$:

$$\text{Expected failures in block } k = \int_0^{T_p} \rho(t)\, dt = \Lambda(T_p),$$

independent of k (because PM resets intensity to $\rho(0)$). Long-run cost:

$$g(T_p) = \frac{c_p + c_f \Lambda(T_p)}{T_p}.$$

Optimal T_p^*: $\rho(T_p^*)T_p^* = \Lambda(T_p^*)$, i.e. the hazard at T_p^* equals the average hazard over $[0, T_p^*]$. For Weibull NHPP ($\rho(t) = \lambda\beta t^{\beta-1}$): $T_p^* = [c_p/(c_f\lambda(\beta-1))]^{1/\beta}$.

Selective maintenance for multi-component systems on missions. Before a mission of duration τ, a subset $\mathcal{S}$ of n components is repaired (with fixed budget B). The reliability of the system during the mission:

$$R(\tau \mid \mathcal{S}) = \phi\big(S_{1|i_1}(\tau), \ldots, S_{n|i_n}(\tau)\big),$$

where $S_{j|0}(\tau)$ is the reliability of an unrepaired component j (accounting for its age), and $S_{j|1}(\tau)$ is the reliability after repair (AGAN or imperfect). The selective maintenance problem: find $\mathcal{S}^* = \arg\max_{|\mathcal{S}| \leq N^*,\, \sum c_i \leq B} R(\tau \mid \mathcal{S})$. Solved by integer programming or greedy heuristic.

Example — Selective maintenance for a naval vessel before patrol

8-component series system. Current ages: $\mathbf{a} = (300, 450, 200, 700, 380, 620, 150, 500)$ hours. Mission duration: 200 h. Budget: 3 repair slots. Weibull components, $k_i = 2$, $\mu_i = 800$ h each.
Mission reliability without maintenance: $R(200) = \prod_i S_i(200 \mid a_i) = \prod_i e^{-(a_i+200)^2/\mu_i^2 + a_i^2/\mu_i^2}$.
Component 4 has highest age (700 h) and contributes most to unreliability. Greedy selection: repair components 4, 6, 2 (highest ages). System mission reliability improves from $R = 0.63$ to $R = 0.89$ with 3 repairs. Comparison with exhaustive search over $\binom{8}{3} = 56$ combinations: greedy solution is optimal here.

Delay-time maintenance model. The delay-time model (Christer, 1982) has two stages: (1) initiation of a defect at rate α; (2) the defect "delays" in a detectable state for time h before

causing failure. Inspections at interval T_I detect defects with probability p_{det}. Expected costs per unit time:

$$g(T_I) = \frac{c_I + c_d \mathbb{E}[\text{defects detected per interval}] + c_f \Pr(\text{failure per interval})}{T_I}.$$

The delay time $h \sim$ Weibull or log-normal, estimated from historical records of inspection findings and failure modes. The optimal T_I^* trades inspection cost against defect-driven failure cost.

(m, T) **replacement policies.** A generalisation of age replacement: replace at planned age T or after the m-th failure within a renewal cycle (whichever comes first), with minimal repair at intermediate failures. This unifies:

- $(1, T)$ policy: replace at first failure or planned age T = standard age replacement.
- (∞, T) policy: replace only at planned age T, with minimal repair at all failures = block replacement with minimal repair.
- (m, ∞) policy: replace after m-th failure, regardless of age = count-based replacement.

The cost rate for the (m, T) policy:

$$g(m, T) = \frac{C_p[1 - G_m(T)] + C_f G_m(T) + c_f \mathbb{E}[\min(N(T), m-1)]}{E[\text{cycle length}]},$$

where $G_m(t) = \Pr(N(t) \geq m)$ is the probability of m or more renewals by time t, and the denominator accounts for the random cycle endpoint.

Failure-limit policies. Replace when the failure rate $h(t)$ exceeds a threshold h^*: $\tau^* = \inf\{t : h(t) \geq h^*\}$. For a Weibull distribution: $\tau^* = h^{*1/(k-1)}/(k\lambda^k)^{1/(k-1)}$. This policy is easy to implement operationally (trigger on hazard estimate) but requires online estimation of $h(t)$ — achieved via the kernel hazard estimator or parametric updating.

Periodic inspection with maintenance decision. Inspections occur at $\Delta, 2\Delta, 3\Delta, \ldots$. At each inspection, the component is either left in service, minimally repaired (cost c_m), or replaced (cost C_p). The decision depends on the observed degradation level $X(k\Delta)$: replace if $X > x_R$, repair if $x_r < X \leq x_R$, do nothing if $X \leq x_r$. The two thresholds (x_r, x_R) and inspection interval Δ are jointly optimised over a 3D grid.

Example — (m, T) policy for offshore pipeline pigging

Offshore oil pipeline: corrosion failures occur as a Poisson process with rate $\lambda = 0.15$/year. "Pigging" (cleaning and inspection) at interval T prevents buildup but does not restore pipe age ($\to$ minimal repair model). Major replacement needed after $m = 3$ pigging interventions that reveal critical corrosion. Costs: $C_p = \$800{,}000$ (section replacement), $c_f = \$50{,}000$ (pigging/minimal repair), $C_f = \$2{,}000{,}000$ (emergency failure repair). Under Poisson: $G_m(T) = 1 - e^{-\lambda T}\sum_{k=0}^{m-1}(\lambda T)^k/k!$.
At $T = 2$ years, $m = 3$: $G_3(2) = 1 - e^{-0.30}(1 + 0.30 + 0.045) = 1 - 0.741 \times 1.345 = 1 - 0.997 = 0.003$. Expected minimal repairs per cycle: $\mathbb{E}[\min(N(2), 2)] = 2G_3(2) + \mathbb{E}[N(2) \wedge 2] \approx 0.006 + \lambda T = 0.30$ (dominated by Poisson). $g(3, 2) \approx (800000 \times 0.997 + 2000000 \times 0.003 + 50000 \times 0.30)/\mathbb{E}[\text{cycle}]$. $\approx (797600 + 6000 + 15000)/2 \approx \$409{,}300$/year. Optimising over $T \in [1, 5]$ and $m \in \{2, 3, 4\}$: minimum at $T^* = 3$ years, $m^* = 4$: $g = \$280{,}000$/year.

Machine learning for maintenance scheduling. Reinforcement learning (RL) formulates maintenance as a Markov Decision Process (MDP): state = observed degradation, action = do nothing, repair, replace, reward = negative cost. The optimal policy $\pi^*(s)$ maximises the expected discounted cumulative reward. Deep Q-networks (DQN) approximate the optimal Q-function $Q(s,a) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t r_t | s_0 = s, a_0 = a, \pi^*]$ via a neural network trained on simulated experience. RL-based maintenance has achieved 15–25% cost reductions over threshold policies in complex multi-component systems, at the cost of interpretability.

7.4 Imperfect Repair Models and Virtual Age

Virtual Age and the Kijima Models

Perfect repair: component is restored to "as good as new" (AGAN) after each repair. Minimal repair: component is restored to "as bad as old" (ABAO) — age is unchanged. *Imperfect repair* interpolates between these extremes.
Kijima Model I: after the k-th failure (at time t_k), the virtual age is $v_k = q \cdot t_k$, where $q \in [0,1]$ is the repair effectiveness. $q = 0$: AGAN. $q = 1$: ABAO.
Kijima Model II: virtual age after k-th repair: $v_k = v_{k-1} + q(t_k - t_{k-1})$, where $t_k - t_{k-1}$ is the k-th inter-failure interval. Each repair partially rejuvenates only the most recent damage.

Hazard under virtual age. Given virtual age v_k after the k-th repair, the next inter-failure time X_{k+1} has hazard:

$$h_{k+1}(t) = h_0(v_k + t), \quad t > 0,$$

where h_0 is the hazard of the underlying lifetime distribution. The repair history is fully encoded in v_k.

Brown–Proschan model. An alternative: after each failure, with probability p repair is perfect (AGAN) and with probability $1-p$ it is minimal (ABAO). This is a special case of random repair effectiveness.

Geometric process. For a deteriorating system, the k-th inter-repair time has $\mathbb{E}[X_k] = a^{k-1}\mu$, where $a > 1$ (decreasing mean — system deteriorates). The geometric process (Lam, 1988) is the standard framework:

$$\{X_k / a^{k-1},\ k = 1, 2, \ldots\} \overset{d}{=} \text{i.i.d. renewal process.}$$

Optimal replacement after N^* repairs: minimise long-run cost rate by solving the integer optimisation over N.

Example — Virtual age model for pump impeller

A pump impeller has underlying Weibull failure time ($k = 2$, $\lambda = 0.005$). After each repair, $q = 0.7$ (Kijima I: system is 70% rejuvenated). Three repairs have occurred at $t_1 = 100$, $t_2 = 180$, $t_3 = 240$ h.
Virtual ages: $v_1 = 0.7 \times 100 = 70$ h; $v_2 = 0.7 \times 180 = 126$ h; $v_3 = 0.7 \times 240 = 168$ h.
Current effective age: 168 h. Hazard at current virtual age: $h_0(168) = 2 \times 0.005 \times (0.005 \times 168)^1 = 0.01 \times 0.84 = 0.0084$ failures/h.
Compare with AGAN (virtual age 0): $h_0(0) = 0$. Compare with ABAO (virtual age = actual age 240 h): $h_0(240) = 0.012$ failures/h.
Decision: at $v_3 = 168$, cost-optimal to replace if $C_p/C_f < S(T_p^* - 168)/[\ldots]$ per the age-replacement formula with modified initial age.

Estimation of q. Given failure history $t_1, t_2, \ldots, t_n$, the partial likelihood for q (with $\boldsymbol{\theta}$ = Weibull parameters) is:

$$L(q, \boldsymbol{\theta}) = \prod_{k=1}^{n} h_0(v_{k-1} + (t_k - t_{k-1})) \cdot S_0(t_k - t_{k-1} \mid v_{k-1}).$$

Profile likelihood over $q \in [0, 1]$ and Weibull parameters simultaneously.

Reliability growth models. As a system undergoes testing and corrective actions, its failure rate decreases over time. The Duane model postulates:

$$h(t) = \frac{\alpha}{\beta} t^{\alpha - 1}, \quad \alpha \in (0, 1),$$

where t is cumulative test time and α is the growth rate. This is the NHPP power law (Crow–AMSAA model). The expected number of failures in $[0, T]$: $\Lambda(T) = (\beta T)^{\alpha}$. MLE: $\hat{\alpha} = n / \ln(T/t_1)$ (for n failures at times $t_1 < \cdots < t_n < T$).

Crow–AMSAA (NHPP Power Law) model.

- Intensity function: $\rho(t) = \lambda \beta t^{\beta - 1}$ (failures/unit time). $\beta < 1$: reliability growth (decreasing intensity). $\beta > 1$: reliability deterioration.
- MLE: $\hat{\beta} = n / \sum_{i=1}^{n} \ln(T/t_i)$; $\hat{\lambda} = n / T^{\hat{\beta}}$.
- Goodness-of-fit: Cramér–von Mises test applied to the transformed times $u_i = (t_i/T)^{\hat{\beta}}$, which should be uniform if the power law holds.

Superimposed NHPP models. For a system where failures are corrected (reducing intensity) but not perfectly (some residual intensity remains), the *generalised Duane–Crow* model uses a Weibull NHPP: $\Lambda(t) = (\lambda t)^{\beta}$. Successive corrections shift the intensity down by amounts $\Delta\rho_i \sim \text{Gamma}(\gamma, \delta)$ — a Bayesian nonparametric reliability growth model.

Pham–Wang Imperfect Repair in NHPP Framework

The Pham–Wang model generalises both minimal repair (NHPP intensity unchanged by repair) and perfect repair (intensity reset to 0): after the k-th repair, the intensity is shifted by $-q_k \rho(t_k^-)$, where $q_k \in [0, 1]$ is the k-th repair effectiveness (possibly random). This gives a piecewise NHPP with changing intensity at each repair epoch.

Example — Crow–AMSAA fit to software reliability growth

A firmware component undergoes 180 days of system testing. Failures at: $t = 2, 5, 9, 18, 30, 55, 89, 130, 165$ days (total $n = 9$, $T = 180$ days). MLE:

$$\hat{\beta} = 9 / \sum_{i=1}^{9} \ln(180/t_i) = 9/(3.85{+}3.58{+}3.00{+}2.30{+}1.79{+}1.19{+}0.70{+}0.32{+}0.09) = 9/16.82 = 0.535.$$

$\hat{\lambda} = 9/180^{0.535} = 9/15.9 = 0.566$.
Since $\hat{\beta} = 0.535 < 1$: reliability is growing. Projected intensity at $T = 180$: $\rho(180) = 0.566 \times 0.535 \times 180^{-0.465} = 0.303 \times 0.060 = 0.018$ failures/day — compared to $\rho(2) = 0.566 \times 0.535 \times 2^{-0.465} = 0.224$ failures/day at the start. An order-of-magnitude improvement. Predicted failures in next 30 days: $\Lambda(210) - \Lambda(180) = 0.566(210^{0.535} - 180^{0.535}) = 0.566 \times (17.0 - 15.9) = 0.62$ failures.

Bayesian update for virtual age parameters. The Bayesian approach places a prior $\pi(q)$ on the repair effectiveness q (e.g. Beta(a, b)). After observing the inter-repair times $\mathbf{T} = (t_1, \dots, t_n)$, the posterior:

$$\pi(q \mid \mathbf{T}) \propto L(\mathbf{T} \mid q)\pi(q),$$

where $L(\mathbf{T} \mid q)$ is the likelihood under the Kijima I model. The posterior mean $\hat{q}^{\text{Bayes}} = \mathbb{E}[q \mid \mathbf{T}]$ and its credible interval quantify uncertainty about repair quality. As n grows, the posterior concentrates around the MLE. For small n (few repair events), the prior plays a significant role — which is appropriate, as physical knowledge about the repair process should inform the estimate.

Self-exciting point processes for repairable systems. When failure history influences future failure rates (e.g. fatigue crack growth, software bugs), the Hawkes process provides a natural model:

$$\rho(t) = \mu + \sum_{t_j < t} g(t - t_j), \quad g(s) = \alpha e^{-\beta s},$$

where each past failure at t_j excites a decaying increase in the failure rate. The Hawkes process generates clustered failures — consistent with fault propagation in complex systems. MLE for (μ, α, β) via the log-likelihood:

$$\ell = \sum_{j=1}^{n} \ln \rho(t_j) - \int_0^T \rho(t)\, dt = \sum_j \ln \rho(t_j) - \mu T - \alpha \sum_j (1 - e^{-\beta(T - t_j)})/\beta.$$

Imperfect repair: statistical estimation challenges. The Kijima model depends on the latent virtual age v_k, which is unobservable. The observed data are only the failure times $\{t_1, \dots, t_n\}$. For fixed repair effectiveness q:

- Virtual ages: $v_1 = qt_1$; $v_k = v_{k-1} + q(t_k - t_{k-1})$ (Kijima II).
- Conditional failure densities: $f_k(t_k - t_{k-1} \mid v_{k-1}) = f_0(v_{k-1} + t_k - t_{k-1})/S_0(v_{k-1})$ (where f_0, S_0 are baseline failure density and survival).

The full likelihood is a product of these conditional densities. When q is unknown, profile likelihood over $q \in [0, 1]$ with Weibull $(\hat{k}, \hat{\lambda})$ is efficient.

Example — Estimating repair quality from recurring compressor failures

A reciprocating compressor experiences failures at: $t = (180, 420, 580, 790, 910, 1080)$ hours. Baseline Weibull with $k = 2.1$, $\mu = 450$ h (from similar virgin units). Profile likelihood over q:
At $q = 0$: virtual ages always reset to 0 (AGAN) — inter-failure times should have mean $\mu = 450$ h. Observed mean inter-failure time: $(180 + 240 + 160 + 210 + 120 + 170)/6 = 180$ h $\ll 450$ h. AGAN inconsistent.
At $q = 1$: virtual ages accumulate (ABAO) — mean inter-failure times should match a renewal process with increasing hazard. Observed mean: 180 h; ABAO predicts decreasing inter-failure times — consistent.
Profile MLE: $\hat{q} = 0.87$ (95% CI: $[0.71, 0.97]$). Repair restores only 13% of wear — consistent with a component that shows fatigue damage accumulation not fully reversed by repair.

Optimal replacement number under geometric process. For a deteriorating system with geometric process parameters (a, λ, k) (where $a > 1$), the long-run cost rate after N repairs

then replace:

$$g(N) = \frac{c_R + \sum_{k=1}^{N} c_k}{\sum_{k=1}^{N} \mu a^{-(k-1)}/[\Gamma(1+1/k_w)\lambda^{-1}]},$$

where $c_k = c_1 q_c^{k-1}$ is the k-th repair cost (escalating at rate q_c) and the denominator is the expected total operating time. Setting $g(N^*) \leq g(N^* \pm 1)$ characterises the optimal replacement number. Closed-form when $c_k = c$ (constant repair cost): $N^* = \lceil \ln(1 - c_R(a-1)/(c\mu_1))/\ln a \rceil$.

7.5 Competing Risks and Multiple Failure Modes

Competing Risks Model

A system with K failure modes has failure time $T = \min(T_1, \ldots, T_K)$ and cause of failure $C = \arg\min_k T_k$. Observable data: (T, C) pairs. Under independence:

$$S_T(t) = \prod_{k=1}^{K} S_k(t), \qquad h_T(t) = \sum_{k=1}^{K} h_k(t),$$

where S_k and h_k are the marginal survival and hazard for cause k.

Cause-specific hazard and subdistribution hazard. Two distinct hazard concepts arise in competing risks:

- **Cause-specific hazard**: $h_k(t) = \lim_{\Delta t \to 0} \Pr(t < T \leq t + \Delta t, C = k \mid T > t)/\Delta t$. Non-parametrically estimated by Nelson–Aalen applied to failures of cause k only (other causes are treated as censoring).
- **Subdistribution hazard** (Fine–Gray): $\tilde{h}_k(t) = -d\ln(1 - F_k(t))/dt$, where $F_k(t) = \Pr(T \leq t, C = k)$ is the *cumulative incidence function* (CIF).

These are *not* equal: $h_k(t)$ governs cause k in isolation; $\tilde{h}_k(t)$ governs the risk of cause k in the presence of all competing causes.

Identifiability. Under competing risks, the joint distribution of $(T_1, \ldots, T_K)$ is *not* identifiable from the marginal data (T, C) without additional assumptions. The marginal survival $S_k(t)$ (latent failure time) is identifiable only under independence (Tsiatis, 1975). This is a fundamental limit: one cannot test independence from competing risks data alone.

Cumulative incidence function. The CIF for cause k is:

$$F_k(t) = \int_0^t S(s^-) h_k(s)\, ds,$$

where $S(s) = \prod_j S_j(s)$ is the overall survival. The Aalen–Johansen estimator is the nonparametric MLE:

$$\hat{F}_k(t) = \sum_{j:\, t_j \leq t} \hat{S}(t_j^-) \frac{d_{kj}}{n_j},$$

where d_{kj} = failures from cause k at t_j.

Example — Competing risks in gas turbine discs

Gas turbine discs can fail by: $k = 1$ (fatigue crack from bore, h_1), $k = 2$ (foreign object damage, h_2), $k = 3$ (creep rupture at high temperature, h_3). Parametric fit to fleet data ($n = 300$ discs):

- Mode 1: Weibull, $\hat{k}_1 = 3.2$, $\hat{\lambda}_1 = 1/15000$ cycles^{-1}.
- Mode 2: Exponential, $\hat{\lambda}_2 = 1.2 \times 10^{-5}$ cycles^{-1} (random events).
- Mode 3: Log-normal, $\hat{\mu}_3 = 9.8$, $\hat{\sigma}_3 = 0.4$ (log-cycles).

Overall survival at $N = 10000$ cycles: $S(10000) = S_1 \times S_2 \times S_3 = e^{-(10000/15000)^{3.2}} \times e^{-0.12} \times \Phi(-1.95) \approx 0.789 \times 0.887 \times 0.974 \approx 0.681$.
CIF for fatigue (mode 1) at 10000 cycles: $\hat{F}_1(10000) \approx 0.18$ — the 18% probability that the first failure is a fatigue crack before 10000 cycles.

Maintenance under competing risks. When cause k is eliminated by a replacement policy (e.g. bore inspection eliminates mode 1 fatigue) but other modes remain, the post-intervention survival is $S^*(t) = S_2(t) \cdot S_3(t)$, which is *not* $S(t)/S_1(t)$ in general (non-identifiability). This shows why lab-derived cause-specific hazards must be used cautiously when translating maintenance interventions to fleet-level predictions.

Part 8

System Reliability and Redundancy

From components to systems: structure functions, redundancy architectures, and importance measures for design and analysis.

Gray zones in competing risks: dependent censoring. If the competing risk C is correlated with T (e.g. withdrawal from study because the patient is deteriorating), treating C as independent censoring in the KM estimator *biases* the survival estimate upward. The bounds on $S_j(t)$ under unidentified dependence are given by the Fréchet–Hoeffding bounds on the joint distribution, which can be wide. Sensitivity analyses perturbing the dependence structure from independence to the Fréchet upper bound provide a range of plausible survival curves.

Fine–Gray subdistribution hazard model. The Fine–Gray regression model:

$$\tilde{h}_k(t \mid \mathbf{x}) = \tilde{h}_{k0}(t) \exp(\tilde{\boldsymbol{\gamma}}_k^\top \mathbf{x}),$$

where $\tilde{h}_k$ is the subdistribution hazard for cause k. Estimation uses a modified risk set: after failure from cause $k' \neq k$, that observation remains in the risk set (with a time-dependent weight) for the analysis of cause k. The estimating equations are weighted partial likelihood equations. *Interpretation*: $\exp(\tilde{\gamma}_{kj})$ is the subdistribution hazard ratio for covariate j on cause k — directly modelling the CIF, accounting for the competing causes.

Cause elimination and latent failure times. A fundamental policy question: "If we could eliminate cause k, how much would overall survival improve?" Under independence: $S_{-k}(t) = \prod_{j \neq k} S_j(t)$. Under dependence: the answer is unidentifiable without additional assumptions. The *net survival* $S_k^{\text{net}}(t)$ (survival in a world where only cause k can kill) is an identifiable quantity only under independence — this is the Tsiatis–Kalbfleisch–Prentice result, a fundamental limitation of observational competing risks data.

Aalen–Johansen Estimator

For K competing causes, the matrix of transition probabilities $\mathbf{P}(s,t)$ (from each transient state at time s to each state at time t) is estimated by:

$$\hat{\mathbf{P}}(s,t) = \prod_{s<t_j\le t} \left(\mathbf{I} + d\hat{\mathbf{A}}(t_j)\right),$$

where $d\hat{\mathbf{A}}(t_j)$ is the matrix of estimated cause-specific hazard increments at t_j: $d\hat{A}_{0k}(t_j) = d_{kj}/n_j$ (from alive-0 to dead-cause-k). The product is ordered. The CIF estimate $\hat{F}_k(t) = \hat{P}_{0k}(0,t)$ is the Aalen–Johansen estimator. It reduces to the KM estimator for $K = 1$.

Example — Competing risks for joint replacement outcomes

$n = 420$ patients receive total hip replacement. Competing events: re-revision (cause 1), death before revision (cause 2). Follow-up: 10 years. Cause-specific hazard estimates:

- Cause 1: decreasing over first 2 years, then stable at ≈ 0.008/year.
- Cause 2: increasing from age-related mortality, ≈ 0.03/year at 10 years.

At $t = 10$ years: $\hat{F}_1(10) = 0.071$ (7.1% probability of re-revision), $\hat{F}_2(10) = 0.152$ (15.2% probability of dying before re-revision). Note $\hat{F}_1 + \hat{F}_2 = 0.223 < 1$ (the remaining 77.7% are alive and un-revised).
Fine–Gray model for cause 1: age at surgery ($\hat{\gamma}_1 = -0.031$/year, $p = 0.041$) and BMI ($\hat{\gamma}_2 = 0.018$/unit, $p = 0.22$). Older patients have lower re-revision risk (likely due to lower activity level), but the effect is significant only after accounting for the competing mortality risk.

Partial likelihood for competing risks. Under K competing causes, the cause-specific Cox partial likelihood for cause k is:

$$\mathcal{L}_k(\boldsymbol{\beta}_k) = \prod_{j:\, C_j = k} \frac{\exp(\boldsymbol{\beta}_k^\top \mathbf{x}_j)}{\sum_{i \in \mathcal{R}(t_j)} \exp(\boldsymbol{\beta}_k^\top \mathbf{x}_i)},$$

treating all failure types $k' \neq k$ as censoring at their failure time. The K likelihoods are maximised separately, giving K independent covariate effect estimates. This approach is computationally simple but the $\boldsymbol{\beta}_k$ estimates do not account for the CIF — for CIF-targeted regression, Fine–Gray is preferred.

Relative survival and net survival. In epidemiology and medical device reliability, the *relative survival* $R_{\text{rel}}(t) = S_{\text{obs}}(t)/S_{\text{expected}}(t)$ compares observed survival to expected survival in a comparable disease-free population. The net survival $S_{\text{net}}(t) = \lim_{n\to\infty} S_{\text{obs}}^{1/n}(t)^n$ (Pohar–Perme estimator) corrects for informative competing causes. This is particularly relevant for implantable devices (pacemakers, stents) where patient age drives both device failure and mortality.

Pseudo-values for competing risks regression. The pseudo-value approach (Andersen–Klein–Rosthøj, 2003): define pseudo-observations for the CIF as $\hat{F}_k^{(-i)}(t)$ (leave-one-out estimator), compute $\hat{V}_k^i(t) = n\hat{F}_k(t) - (n-1)\hat{F}_k^{(-i)}(t)$, and fit a regression:

$$g(\mathbb{E}[V_k^i(t) \mid \mathbf{x}_i]) = \boldsymbol{\beta}_k^\top \mathbf{x}_i$$

using GEE (Generalised Estimating Equations) with the logit or complementary log-log link. This gives direct regression on the CIF without the proportionality constraint of Fine–Gray.

Latent class models for heterogeneous failure populations. When the population consists of C latent subgroups with different cause-specific hazards, a latent class competing risks model:

$$f(t, k \mid \mathbf{x}) = \sum_{c=1}^{C} \pi_c(\mathbf{x})\, h_{kc}(t)\, S_c(t),$$

where $\pi_c(\mathbf{x}) = \exp(\boldsymbol{\alpha}_c^\top \mathbf{x}) / \sum_{c'} \exp(\boldsymbol{\alpha}_{c'}^\top \mathbf{x})$ (multinomial logit class probability). Estimation via EM: E-step computes the posterior class probability for each observation; M-step updates cause-specific hazards and class probabilities. Identifiability requires either parametric hazards or informative covariates.

Example — Latent class model for medical device failures

$n = 620$ cardiovascular stent implants followed for 10 years. Two competing events: stent thrombosis ($k = 1$) and restenosis ($k = 2$). Hypothesis: two latent patient subpopulations — high-risk ($C = 1$, 25%) and low-risk ($C = 2$, 75%). Fitted model:

- High-risk ($\hat{\pi}_1 = 0.23$): elevated thrombosis hazard ($\hat{\lambda}_{11} = 0.08$/year), moderate restenosis ($\hat{\lambda}_{21} = 0.12$/year).
- Low-risk ($\hat{\pi}_2 = 0.77$): low thrombosis ($\hat{\lambda}_{12} = 0.01$/year), low restenosis ($\hat{\lambda}_{22} = 0.04$/year).

The latent class model fits the bimodal failure pattern (early failures followed by a long tail) better than a homogeneous competing risks model (AIC: 1842 vs. 1908).

Multi-state models for device degradation. A multi-state model extends competing risks by allowing intermediate states (e.g. device degraded but functional, device requires adjustment, device failed). The system evolves as a CTMC or semi-Markov process, with transition intensities estimated by the Aalen–Johansen estimator. The *sojourn time* in each state and the *length-biased sampling* effect (devices more often observed in longer-lived states) require the Bayes–bootstrap length-bias correction.

Deep Dive: Point Process Models for Repairable Systems

General counting process framework. Let $N(t)$ count the number of failures in $[0, t]$. The *intensity function* $\lambda(t) = \lim_{\Delta t \to 0} \mathbb{E}[\Delta N(t) | \mathcal{F}_{t^-}] / \Delta t$ generalises the hazard function to the case of repeated events. For:

- Perfect repair: $\lambda(t) = h(t - S_{N(t^-)})$ (hazard from last renewal time) — ordinary renewal process.
- Minimal repair: $\lambda(t) = h(t)$ (hazard of original distribution evaluated at actual age) — NHPP.
- Imperfect repair: $\lambda(t) = h(t - \tilde{S}_{N(t^-)})$ where $\tilde{S}$ is the virtual age — Kijima model.

Log-likelihood for general repairable systems. Given failure times $t_1 < t_2 < \cdots < t_n$ in $[0, T]$:

$$\ell = \sum_{i=1}^{n} \ln \lambda(t_i) - \int_0^T \lambda(t)\, dt.$$

For the power law NHPP ($\lambda(t) = \lambda\beta t^{\beta-1}$): $\ell = n \ln(\lambda\beta) + (\beta - 1) \sum_i \ln t_i - \lambda T^\beta$.

Residuals for point process models. The *conditional intensity residuals*: $\hat{e}_i = \int_{t_{i-1}}^{t_i} \hat{\lambda}(t)\,dt$ (expected events between failures). Under a correct model: $\hat{e}_i \overset{iid}{\sim} \text{Exp}(1)$ (by the random time change theorem). Plot $\hat{e}_i$ vs. i: systematic trends indicate model misspecification. The QQ plot of $\hat{e}_i$ vs. $\text{Exp}(1)$ quantiles is the standard diagnostic.

Trend tests. Before fitting a specific repairable system model, test for trend in the inter-failure times:

- *Laplace test*: $U = (\bar{t} - T/2)/\sqrt{T^2/(12n)} \sim \mathcal{N}(0,1)$ under H_0 (HPP). $U > 0$: reliability decreasing; $U < 0$: reliability improving.
- *Military Handbook test*: $\chi^2 = 2n/[\beta T^\beta] \sum_i T_{(i)}^\beta$ under the power law HPP — tests $\hat{\beta} = 1$ vs. $\hat{\beta} \neq 1$.
- *Lewis–Robinson test*: tests for independent increments vs. positive trend, robust to non-Weibull alternatives.

Example — Trend test and NHPP fit for a diesel generator

A standby diesel generator experiences 11 failures over 5 years. Failure times (days): 45, 128, 210, 295, 380, 445, 510, 620, 710, 820, 910.
Laplace test: $\bar{t} = (45 + 128 + \cdots + 910)/11 = 465$ days. $T = 1825$ days (5 years). $U = (465 - 912.5)/\sqrt{1825^2/(12 \times 11)} = -447.5/198.6 = -2.25$. $|U| > 1.96$: statistically significant *decreasing* failure rate ($p = 0.025$) — reliability growth is occurring (maintenance improvements reducing failures).
Power law NHPP: $\hat{\beta} = 11/\sum_i \ln(1825/t_i) = 11/(3.70 + 2.65 + 2.16 + 1.83 + 1.57 + 1.41 + 1.27 + 1.08 + 0.94 + 0.80 + 0.70) = 11/17.11 = 0.643$. $\hat{\beta} < 1$ confirms reliability growth. Predicted failures in year 6: $\hat{\lambda}(1825 \to 2190) = \hat{\lambda}[(2190/1825)^{0.643} - 1] \times 1825^{0.643}/2190^{0.643} \approx 1.6$ additional failures (vs. 2.2/year if no growth).

8.1 Structure Functions: Series, Parallel, and k-out-of-n Systems

Structure Function

A system of n binary components has state vector $\mathbf{x} = (x_1, \ldots, x_n) \in \{0,1\}^n$, where $x_i = 1$ (working) or $x_i = 0$ (failed). The *structure function* $\phi : \{0,1\}^n \to \{0,1\}$ maps component states to system state ($\phi = 1$: working; $\phi = 0$: failed).

- **Series**: $\phi(\mathbf{x}) = \prod_{i=1}^n x_i = \min_i x_i$. All components must work.
- **Parallel**: $\phi(\mathbf{x}) = 1 - \prod_{i=1}^n (1 - x_i) = \max_i x_i$. At least one must work.
- **k-out-of-n**: $\phi(\mathbf{x}) = \mathbf{1}\left[\sum_{i=1}^n x_i \geq k\right]$.

Reliability polynomial. Under component independence with $p_i = \Pr(x_i = 1)$, the system reliability is:

$$R_s = \mathbb{E}[\phi(\mathbf{x})] = \sum_{\mathbf{x} \in \{0,1\}^n} \phi(\mathbf{x}) \prod_{i=1}^n p_i^{x_i} (1 - p_i)^{1 - x_i}.$$

For series: $R_s = \prod_i p_i$. For parallel: $R_s = 1 - \prod_i (1 - p_i)$. For k-out-of-n with $p_i = p$:

$$R_{k/n} = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}.$$

Coherent systems. A structure function is *coherent* if: (i) it is non-decreasing in each x_i (replacing a failed component cannot hurt), and (ii) $\phi(\mathbf{0}) = 0$ and $\phi(\mathbf{1}) = 1$. All practical reliability structures are coherent.

Minimal cut sets and path sets. A *minimal path set* $P \subseteq \{1, \dots, n\}$ is a minimal set of components whose simultaneous working guarantees system function: $\phi(\mathbf{x}) = 1$ if $x_i = 1 \, \forall \, i \in P$. A *minimal cut set* K is a minimal set whose simultaneous failure causes system failure. Then:

$$\phi(\mathbf{x}) = \max_{P \in \mathcal{P}} \prod_{i \in P} x_i = \min_{K \in \mathcal{K}} \max_{i \in K} x_i,$$

giving a series-of-parallel (path representation) and a parallel-of-series (cut representation) respectively.

Example — Bridge network structure function

Five components arranged as a bridge: 1–2 (top path), 3–4 (bottom path), 5 (bridge). Minimal path sets: $\{1,2\}$, $\{3,4\}$, $\{1,5,4\}$, $\{2,5,3\}$. Minimal cut sets: $\{1,3\}$, $\{2,4\}$, $\{1,5,4\}$... wait, the minimal cut sets are $\{1,3\}$, $\{2,4\}$, $\{1,4,5\}$... For equal $p_i = 0.9$: $R_s = \Pr(\text{at least one path works}) = 1 - \Pr(\text{all cuts fail})$. Via inclusion-exclusion on cuts:

$$R_s = 1 - [(1-p^2)^2 - \dots] \approx 0.9785 \text{ (exact: } 2p^2 + 2p^3 - 5p^4 + 2p^5).$$

At $p = 0.9$: $R_s = 2(0.81) + 2(0.729) - 5(0.6561) + 2(0.59049) = 1.62 + 1.458 - 3.2805 + 1.18098 = 0.97848$.

Time-dependent reliability. For exponentially failing components with rates λ_i:

$$R_s^{\text{series}}(t) = \exp\left(-\sum_i \lambda_i t\right) = \exp(-\lambda_{\text{sys}} t), \quad \lambda_{\text{sys}} = \sum_i \lambda_i.$$
$$R_s^{\text{parallel}}(t) = 1 - \prod_i (1 - e^{-\lambda_i t}).$$

For a parallel system, the MTTF is $\text{MTTF} = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k}^{-1} \sum_{|S|=k} \left(\sum_{i \in S} \lambda_i\right)^{-1}$.

Reliability bounds for complex systems. Computing exact system reliability requires enumeration of 2^n states (exponential in n). Upper and lower bounds provide computationally tractable alternatives:

Inclusion-exclusion bounds (Bonferroni).

$$R_s \le \sum_{P \in \mathcal{P}} \Pr\Big(\bigcap_{i \in P} \{x_i = 1\}\Big) \quad \text{(first-order upper bound)},$$
$$R_s \ge \sum_{P} \Pr\Big(\bigcap_{i \in P} \{x_i = 1\}\Big) - \sum_{P < Q} \Pr\Big(\bigcap_{i \in P \cup Q} \{x_i = 1\}\Big) \quad \text{(second-order lower bound)}.$$

Abraham (1979) bound. For minimal path sets:

$$R_s \ge 1 - \prod_{P \in \mathcal{P}} (1 - \Pr(\text{path } P \text{ works})),$$

assuming path failures are independent — an approximation that gives a lower bound under positive dependence.

Inclusion-exclusion exact formula. For m minimal path sets:

$$R_s = \sum_{k=1}^{m} (-1)^{k+1} \sum_{|I|=k} \Pr\Big(\bigcap_{j \in I} A_j\Big),$$

where $A_j = \{\text{path } P_j \text{ works}\} = \bigcap_{i \in P_j} \{x_i = 1\}$. Under independence: $\Pr(\bigcap_{j \in I} A_j) = \prod_{i \in \bigcup_{j \in I} P_j} p_i$.

ESREL (event-space reliability) via BDD. Binary Decision Diagrams (BDD) represent the structure function as a directed acyclic graph, enabling exact reliability computation in $O(2^n)$ worst case but $O(\text{poly}(n))$ for many practical topologies (series-parallel, k-out-of-n, modular). Variable ordering heuristics (minimum bandwidth, force-directed) reduce BDD size by orders of magnitude.

Example — BDD for a 5-component bridge

The bridge structure function $\phi = x_1x_2 + x_3x_4 + x_1x_5x_4 + x_2x_5x_3$ (sum of minimal paths, with overlaps resolved by inclusion-exclusion) is represented as a BDD with root variable x_1. With variable order x_1, x_3, x_5, x_2, x_4:

$$R_s = p_1[p_2+(1-p_2)(p_3p_5+p_3(1-p_5)p_4+(1-p_3)p_4p_5)]+(1-p_1)[p_3p_4+p_3(1-p_4)p_2p_5+(1-p_3)p_4p_2p_5].$$

At $p_i = 0.95$: $R_s = 0.9785$ (matching inclusion-exclusion). BDD evaluates in 31 operations vs. $2^5 = 32$ state enumerations — marginal here but dramatic for $n = 50$ where BDD can have $O(n^2)$ nodes.

Multi-state reliability. In multi-state systems (MSS), component state $X_i \in \{0, 1, \dots, m_i\}$ represents performance levels (e.g. 0 = failed, 1 = degraded, 2 = full capacity). System performance $\Phi(\mathbf{X}) \in \{0, 1, \dots, M\}$. Reliability is defined as $R(d) = \Pr(\Phi \geq d)$ for demand d. The universal generating function (UGF) method:

$$U_i(z) = \sum_{k=0}^{m_i} p_{ik} z^{g_{ik}},$$

where g_{ik} is the performance of state k. For a series system: $U_s(z) = \otimes_i U_i(z)$ using the composition operator $\otimes$ defined by min of performance: $(p_a z^{g_a}) \otimes (p_b z^{g_b}) = p_a p_b z^{\min(g_a, g_b)}$. Reliability: $R(d) = \Pr(g \geq d) = \sum_{g \geq d} p_g$ extracted from $U_s(z)$.

Modules and modular decomposition. A set $M \subset [n]$ is a *module* of a structure function ϕ if $\phi(\mathbf{x})$ depends on $\{x_i : i \in M\}$ only through $\phi_M(\mathbf{x}_M)$ (the sub-structure function of M). Modular decomposition recursively identifies a tree of modules, enabling divide-and-conquer reliability computation: compute the reliability of each module separately, then assemble. For series-parallel systems, the entire tree is found in $O(n \log n)$ time.

Probabilistic risk assessment (PRA) for nuclear power plants. Full-scope Level 1 PRA integrates:

1. *Initiating event analysis*: identify and quantify all initiating events (LOCA, loss of feedwater, etc.) with frequencies λ_{IE}.
2. *Event tree analysis*: develop success/failure scenarios for each initiating event, quantifying sequences by multiplying branch probabilities.
3. *Fault tree analysis*: quantify each top event (system failure probability) feeding into the event tree.
4. *Large early release frequency (LERF)*: integrate over all scenarios with severe core damage to estimate annual LERF.

Uncertainty analysis in PRA. Each basic event probability p_i is uncertain, represented by a

lognormal distribution $p_i \sim \text{LN}(\mu_i, \sigma_i)$ (epistemic uncertainty). Two-loop Monte Carlo:

- Outer loop: sample $p_i^{(k)}$ from their epistemic distributions.
- Inner loop (optional): compute point estimate of P_f given $\{p_i^{(k)}\}$.

The result is a distribution of P_f — typically log-normal with geometric mean (point estimate) and 5th/95th percentiles (uncertainty bounds). NUREG-1150 targets: mean CDF $< 10^{-4}$/reactor-year, 95th percentile $< 10^{-3}$/reactor-year.

Example — Cut set importance in a nuclear safety system

Level 1 PRA for a simplified emergency core cooling system (ECCS). After fault tree analysis: 12 minimal cut sets. Top 5 by Fussell–Vesely importance:

Cut set	$P(K_j)$	FV importance
{Pump A, Pump B}	4.0×10^{-5}	0.31
{Common instrumentation}	3.2×10^{-5}	0.25
{Valve V1, Pump A}	1.8×10^{-5}	0.14
{Valve V1, Pump B}	1.8×10^{-5}	0.14
{Power supply}	1.5×10^{-5}	0.12

Total system failure probability $\approx 1.27 \times 10^{-4}$/demand. The common instrumentation cut set (single basic event) is a single-point vulnerability — addressed by adding a diverse (pneumatic) instrumentation channel.

Hierarchical reliability models. For large complex systems (aircraft, power grids), reliability is computed hierarchically: subsystem reliabilities are computed from component-level FTA/RBD; system reliability is computed from subsystem reliabilities using higher-level FTA. Uncertainty propagates through the hierarchy: $\sigma^2_{\ln R_s} \approx \sum_j (dR_s/dR_j)^2 \sigma^2_{\ln R_j}$ (FOSM), or exactly via nested Monte Carlo. The hierarchical decomposition enables parallel computation across engineering teams.

8.2 Reliability Block Diagrams and Fault Trees

RBD and FTA

A *Reliability Block Diagram* (RBD) represents component connections graphically: blocks in series = series system; blocks in parallel = parallel subsystem. More complex systems are represented as combinations.
A *Fault Tree* (FT) is a top-down deductive model: the *top event* (system failure) is decomposed via logic gates (AND, OR, K-of-N) into intermediate events and ultimately to *basic events* (component failures).

Qualitative analysis of fault trees. The *minimal cut sets* of the fault tree are found by Boolean algebra reduction of the tree. For an OR gate, the MCS of the output is the union of MCS of inputs. For an AND gate, the MCS of the output is the Cartesian product of MCS of inputs (pair-wise combinations).

Quantitative FT analysis. Under independent basic event probabilities p_i:

$$P(\text{top event}) = P\left(\bigcup_{j=1}^{m} K_j\right) \approx \sum_{j=1}^{m} P(K_j) - \sum_{j<k} P(K_j \cap K_k) + \cdots$$

The inclusion-exclusion series is exact but $O(2^m)$; the *rare event approximation* $P \approx \sum_j \prod_{i \in K_j} p_i$ is $O(m)$ and accurate when $p_i \ll 1$.

Common cause failures (CCF) in fault trees. CCF groups connect multiple components through a common-cause basic event (e.g. a shared environment). The β-factor model: each component has independent failure probability $q_I = (1-\beta)q$ and shares a common-cause event of probability $q_C = \beta q$, so $q = q_I + q_C$.

Example — Fault tree for a redundant pumping system

Top event: Loss of Flow (LOF). The system has two pumps (P1, P2) in parallel, each requiring an electrical supply (E1, E2). The tree:

- LOF = AND(P1 fails OR E1 fails, P2 fails OR E2 fails) [both subsystems must fail].
- MCS: {P1, P2}, {P1, E2}, {E1, P2}, {E1, E2}.

With $p_{\text{pump}} = 10^{-3}$, $p_{\text{elec}} = 5 \times 10^{-4}$:

$$P(\text{LOF}) \approx (10^{-3})^2 + 10^{-3}\times 5\times 10^{-4} + 5\times 10^{-4}\times 10^{-3} + (5\times 10^{-4})^2 = 10^{-6} + 10^{-6} + 2.25\times 10^{-7} = 2.225\times 10^{-6}.$$

With $\beta = 0.1$ CCF on pumps: additional CCF event $q_C = 0.1 \times 10^{-3} = 10^{-4}$. New $P(\text{LOF}) \approx 2.225 \times 10^{-6} + 10^{-4} \times 5 \times 10^{-4} + 10^{-4} \times 5 \times 10^{-4} \approx 1.02 \times 10^{-4}$ — a 50× increase!

Event tree analysis (ETA). ETA models sequences of events following an initiating event, quantifying scenarios by multiplying branch probabilities. FTA and ETA are complementary: FTA identifies what causes the initiating event; ETA traces its consequences. Their combination is the *bow-tie model* widely used in process safety.

Dynamic fault trees (DFT). Standard fault trees use only static AND/OR gates, capturing instantaneous logic but not dynamic behaviour (repair, sequencing, spare switching). DFT adds:

- **Priority AND (PAND)**: output fails iff all inputs fail in left-to-right sequence order.
- **Spare gate**: output fails iff primary and all spare components have failed; spares switch in when primary fails (warm/cold spare modes).
- **Sequence enforcing (SEQ)**: forces failures to occur in a specified order.
- **Functional dependency (FDEP)**: a trigger event forces a set of components to fail simultaneously.

DFTs are analysed via modular decomposition (static subtrees use standard FT methods; dynamic modules use Markov chains) or directly by conversion to CTMCs.

Markov chain model for repairable systems. For n binary components each with failure rate λ_i and repair rate μ_i, the state space is $\{0,1\}^n$ (size 2^n). The CTMC transition rates: $q_{s \to s'} = \lambda_i$ if $s' = s$ with component i flipped from 1 to 0 (failure), and μ_i if flipped from 0 to 1 (repair). System unavailability: $U(t) = \sum_{s:\,\phi(s)=0} \pi_s(t)$. Asymptotic (steady-state): solve $\boldsymbol{\pi} Q = \mathbf{0}$, $\boldsymbol{\pi}\mathbf{1} = 1$.

For large n, the CTMC has 2^n states — intractable by direct methods. Approximations: (1) *modular decomposition* (Section 8.2 of DFT); (2) *Boolean network reduction* (group states by relevance); (3) *simulation* (Monte Carlo CTMC sampling).

Example — Markov model for a 2-out-of-3 repairable system

Three identical components, $\lambda = 0.01$ f/h, $\mu = 0.1$ r/h. States: $\{3, 2, 1, 0\}$ = number of working components. Transitions: $3 \to 2$: rate $3\lambda = 0.03$; $2 \to 1$: rate $2\lambda = 0.02$; $1 \to 0$: rate $\lambda = 0.01$; $2 \to 3$: rate $\mu = 0.1$; $1 \to 2$: rate $\mu = 0.1$; $0 \to 1$: rate $\mu = 0.1$ (one repair at a time).
Steady-state (solving $\pi Q = 0$):

$$\pi_3 = 0.9694,\ \pi_2 = 0.02909,\ \pi_1 = 0.001455,\ \pi_0 = 0.0000727.$$

System unavailability (fails in states 1 and 0 under 2/3 logic): $U = \pi_1 + \pi_0 \approx 0.001528$. Mean time between system failures (via first-passage): MTBF $\approx 1/(U \cdot \lambda_{\text{sys}})$, where $\lambda_{\text{sys}} = q_{2\to1} = 0.02$ per unit time: MTBF $\approx 1/(0.001528 \times 0.02)^{-1} \approx 32700$ h.

Petri nets for reliability. Stochastic Petri nets (SPN) model system dynamics via tokens in places, transitions firing at exponential rates. Places represent component states; transitions represent failures, repairs, or demand events. The marking reachability graph is equivalent to the CTMC; SPN software (PIPE, GreatSPN) generates and solves the CTMC automatically from the graphical model. Generalised SPN (GSPN) allows immediate (zero-time) transitions for logical conditions.

Phased mission analysis. A system must complete K phases (e.g. a spacecraft launch, cruise, and landing). Failure in any phase ends the mission. Phase k lasts duration T_k with its own structure function ϕ_k and component failure rates $\lambda_i^{(k)}$. Mission reliability:

$$R_{\text{mission}} = \prod_{k=1}^{K} \Pr(\text{system survives phase } k \,|\, \text{survived phases } 1, \ldots, k-1).$$

The conditional probability must account for the age of components entering each phase. The *phased-mission BDD* (Zang et al., 1999) computes this exactly by expanding component variables per phase.

Software reliability growth models. Software failures are non-physical but follow reliability growth patterns. The Jelinski–Moranda (JM) model: n_0 initial faults, each with equal removal probability. After removing i faults, the failure rate is $\rho_i = \phi(n_0 - i)$. MLE for (n_0, ϕ) from failure times:

$$\hat{\phi} = k / \left[\hat{n}_0 \sum_{j=1}^{k} t_j - \sum_{j=1}^{k} (j-1) t_j \right],$$

where k is the number of observed failures and t_j are times between failures. The Musa–Okumoto logarithmic Poisson model: $\Lambda(t) = \lambda_0 \ln(1 + \lambda_1 t)$ — failure rate decreases as $\lambda(t) = \lambda_0 \lambda_1 / (1 + \lambda_1 t)$. Current failures per unit time decrease as testing proceeds; predicted residual failures: $N_\infty - N(t) =$ remaining bugs.

Physics of failure (PoF) models. PoF integrates physics-based failure mechanisms directly into reliability prediction, replacing empirical Weibull fits with mechanistic models:

- *Electromigration* in IC metal lines: MTTF $= Aj^{-n} \exp(E_a/kT)$ (Black's equation), where j is current density and $E_a \approx 0.7$ eV for Al.
- *Dielectric breakdown*: time-dependent dielectric breakdown (TDDB) with MTTF $\propto \exp(B/E_{\text{ox}})$ (E-model) or $\exp(\gamma E_{\text{ox}})$ (E-model variant).
- *Corrosion*: Butler–Volmer electrochemical kinetics, Tafel equation for anodic current.

- *Fatigue*: Paris–Erdogan (Section 6.3), Coffin–Manson for solder joints.

Example — MTBF prediction for a circuit board via JEDEC models

A circuit board contains: 2 BGAs ($\lambda_{\text{BGA}} = 2.5 \times 10^{-9}$/h), 12 decoupling capacitors ($\lambda_{\text{cap}} = 5 \times 10^{-10}$/h), and a microcontroller ($\lambda_{\mu C} = 3.5 \times 10^{-8}$/h). Operating at 70°C (vs. 25°C reference); Arrhenius with $E_a = 0.6$ eV:

$$AF = \exp\left(\frac{0.6}{8.617 \times 10^{-5}}\left(\frac{1}{298} - \frac{1}{343}\right)\right) = \exp(9.07) \approx 8700.$$

Adjusted failure rates at 70°C: $\lambda_{\text{BGA}}^{70^\circ} = 2.5 \times 10^{-9} \times 8700 = 2.2 \times 10^{-5}$/h.
System failure rate: $\lambda_s = 2 \times 2.2 \times 10^{-5} + 12 \times 4.4 \times 10^{-6} + 3.0 \times 10^{-4} = 4.4 \times 10^{-5} + 5.3 \times 10^{-5} + 3.0 \times 10^{-4} \approx 3.9 \times 10^{-4}$/h. MTBF $= 1/\lambda_s \approx 2560$ h at 70°C.

Functional failure analysis (FMEA/FMECA). Failure Mode and Effects Analysis systematically lists all component failure modes, their effects on subsystem and system function, occurrence probability, severity, and detectability. The Risk Priority Number (RPN):

$$\text{RPN} = \text{Severity} \times \text{Occurrence} \times \text{Detection}$$

(each scored 1–10) ranks failure modes for mitigation priority. FMECA (Criticality Analysis) extends FMEA by computing criticality $C_m = \beta\alpha\lambda_p t$, where β is conditional probability that the failure mode causes the end effect, α is mode ratio, λ_p is part failure rate, and t is operating time.

Fault tree construction methodology. The construction of a correct fault tree requires a disciplined top-down approach:

1. **Define the top event precisely**: be specific about the system boundary, failure mode, and operating conditions (e.g. "Loss of cooling water flow to reactor during normal operation" not "cooling failure").
2. **Identify the immediate necessary causes**: use an AND gate if all causes must be present simultaneously; OR gate if any one suffices.
3. **Develop each intermediate event**: continue down until reaching basic events (component failures with known probabilities from databases or tests).
4. **Mark transfer gates**: identical sub-trees are referenced by transfer symbols to avoid duplication.
5. **Validate**: walk through every accident scenario to verify the tree correctly captures all failure paths.

Inhibit gate and conditional events. The inhibit gate outputs a failure when: (a) the input event occurs AND (b) a conditional event (usually a non-random enabling condition, e.g. maintenance personnel present) is true. The inhibit gate is used for human error events where the human error occurs only if the operator is performing a specific task.

NOT gate and success trees. By De Morgan's theorem, a fault tree can be converted to a *success tree* by replacing AND gates with OR, OR gates with AND, and complementing all basic events. The success tree models the conditions for system success. Success trees are used to verify that at least one success path (minimal path set) exists and is achievable.

Quantitative FTA: importance of ordering. When basic event probabilities span many orders of magnitude, the inclusion-exclusion series for P_f converges slowly and can be domi-

nated by numerical cancellation. The preferred algorithm for exact computation is:

1. Find all MCS of size $\leq K_{\max}$ (truncated expansion).
2. Use BDD for exact computation of larger trees.
3. For ultra-rare events ($P_f < 10^{-10}$): use the binary inclusion-exclusion on MCS sorted by probability.

Example — Fault tree for a chemical reactor emergency shutdown system

Top event: "Reactor temperature exceeds $T_{\max}$ for more than 5 seconds" (initiates damage).
Level 1 (OR): Cooling failure OR Temperature sensor failure AND control system failure.
Level 2:

- Cooling failure (AND): pump failure AND emergency coolant valve fails to open.
- Control system failure (AND): primary controller fails AND backup controller fails.

Basic event probabilities (per demand): pump failure $= 10^{-3}$, valve $= 5 \times 10^{-3}$, primary controller $= 2 \times 10^{-3}$, backup controller $= 10^{-3}$, temp. sensor $= 5 \times 10^{-4}$.
MCS: {pump, valve} with $P = 10^{-3} \times 5 \times 10^{-3} = 5 \times 10^{-6}$; {sensor, primary, backup} with $P = 5 \times 10^{-4} \times 2 \times 10^{-3} \times 10^{-3} = 10^{-9}$.
$P(\text{top}) \approx 5 \times 10^{-6} + 10^{-9} \approx 5 \times 10^{-6}$. The cooling failure path dominates by three orders of magnitude. Priority: reduce pump or valve failure probability.

Human reliability analysis (HRA) integration. Human errors are modelled as basic events in the fault tree with probability given by HRA techniques:

- **THERP** (Technique for Human Error Rate Prediction): tabulated error probabilities for specific action types, combined with performance shaping factors (PSF: time pressure, stress, procedure quality).
- **ATHEANA** (A Technique for Human Event Analysis): focuses on error-forcing context — the combination of plant conditions that makes human error highly likely.
- **SPAR-H**: simplified version of THERP used in NRC safety analyses; 8 PSF multipliers on nominal HEP.

Human error typically contributes 30–70% of total core damage frequency in nuclear PRA, making HRA one of the most important components of the analysis.

Uncertainty quantification in FTA. Basic event probabilities are uncertain (epistemic uncertainty from limited data). Two approaches:

- *Monte Carlo propagation*: sample $(p_1, \ldots, p_n)$ from their uncertainty distributions (typically log-normal with given error factor EF: 5th/95th percentile ratio $= \text{EF}^2$), compute P_f for each sample. Output: distribution of P_f.
- *Analytical moment propagation*: for the rare-event approximation $P_f \approx \sum_j P(K_j)$, and log-normal basic events: P_f is approximately log-normal with $\mu_{\ln P_f} = \sum_j \mu_{\ln P(K_j)}$ and $\sigma^2_{\ln P_f} = \sum_j \sigma^2_{\ln P(K_j)}$ (if MCS are independent).

8.3 Importance Measures: Birnbaum, Criticality, and Fussell–Vesely

Component Importance Measures

Let $R_s = \mathbb{E}[\phi(\mathbf{x})]$ be the system reliability and $p_i = \Pr(x_i = 1)$. Three standard measures:

Birnbaum importance: $I_i^B = \dfrac{\partial R_s}{\partial p_i} = R_s(1_i, \mathbf{p}) - R_s(0_i, \mathbf{p})$, where $R_s(1_i, \mathbf{p})$ means component i is definitely working.

Criticality importance: $I_i^C = I_i^B \cdot \frac{(1-p_i)}{1-R_s}$ (relative contribution of component i's failure to system failure).

Fussell–Vesely importance: $I_i^{FV} = \dfrac{\sum_{K_j \ni i} P(K_j)}{1-R_s}$ (fraction of system unreliability attributable to MCS containing component i).

Birnbaum interpretation. I_i^B is the probability that component i is *critical* — i.e., that the system is in a state where component i's failure or repair determines the system state. It is the marginal increase in system reliability per unit increase in p_i.

Risk Achievement Worth (RAW) and Risk Reduction Worth (RRW).

$$\text{RAW}_i = \frac{1 - R_s(0_i, \mathbf{p})}{1 - R_s}, \qquad \text{RRW}_i = \frac{1 - R_s}{1 - R_s(1_i, \mathbf{p})}.$$

RAW_i: factor by which system unreliability increases if component i is always failed (worst case). RRW_i: factor by which system unreliability decreases if component i never fails (best case). Both are widely used in nuclear safety (NUREG-1150).

Structural importance. $I_i^S = I_i^B\big|_{p_j = 0.5\ \forall j}$ — the Birnbaum importance evaluated at equal component reliabilities. It depends only on the structure function, not the p_i values, and ranks components by their topological centrality.

Example — Importance measures for a 2-out-of-3 system

Three identical components, $p_i = p = 0.9$. k-out-of-$n = 2/3$ system.
$R_s = 3p^2(1-p) + p^3 = 3(0.81)(0.1) + 0.729 = 0.243 + 0.729 = 0.972$.
$R_s(1_i) = p^2 + 2p(1-p) = 0.81 + 0.18 = 0.99$ (with i working, need at least one of the other two).
Wait: with component i fixed at 1, system works if at least 1 of 2 others work: $R_s(1_i) = 1 - (1-p)^2 = 1 - 0.01 = 0.99$.
$R_s(0_i) = p^2 = 0.81$ (both remaining must work).
$I_i^B = 0.99 - 0.81 = 0.18$.
$I_i^C = 0.18 \times (1-0.9)/(1-0.972) = 0.18 \times 0.1/0.028 \approx 0.643$.
I_i^{FV}: MCS are $\{1,2\}, \{1,3\}, \{2,3\}$. Each $P(K_j) = (0.1)^2 = 0.01$. Component i appears in 2 MCS: $I_i^{FV} = 2 \times 0.01/0.028 \approx 0.714$.
By symmetry all components have equal importance; the system is equally sensitive to each component.

Time-dependent importance. For age-dependent component reliabilities $p_i(t)$, all importance measures become functions of t. The *lifetime importance* integrates over the mission time:

$$\bar{I}_i = \frac{1}{t_m} \int_0^{t_m} I_i(t)\, dt.$$

Differential importance measure (DIM). The DIM extends Birnbaum importance to account for the rate at which each component's reliability is being improved:

$$I_i^{DIM} = \frac{I_i^B \cdot \Delta p_i}{\sum_j I_j^B \cdot \Delta p_j},$$

where Δp_i is the planned improvement in component i's reliability (e.g. from a design change). The DIM ranks components by their expected contribution to system reliability improvement under the given improvement plan.

Joint importance. When two components' reliabilities are simultaneously changed:

$$I_{ij}^J = \frac{\partial^2 R_s}{\partial p_i \partial p_j} = R_s(1_i, 1_j) - R_s(1_i, 0_j) - R_s(0_i, 1_j) + R_s(0_i, 0_j).$$

$I_{ij}^J > 0$: components i and j are *jointly critical* (they share minimal cut sets). $I_{ij}^J < 0$: components are *redundant* (knowing one works reduces the importance of the other).

Temporal importance. For mission reliability over a time horizon $[0, T]$:

$$I_i^B(t) = \frac{\partial R_s(t)}{\partial R_i(t)},$$

where $R_i(t) = S_i(t) = e^{-\lambda_i t}$ for exponential components. The time-average importance $\bar{I}_i = T^{-1} \int_0^T I_i^B(t)\, dt$ weights early vs. late failures according to their impact. For series systems: $\bar{I}_i = 1$ for all i (each component is equally structurally important). For parallel systems: $\bar{I}_i$ decays with time as other components provide redundancy.

Reliability Importance Ranking for Maintenance Prioritisation

Given limited maintenance budget B and a set of potential upgrades $(i, \Delta R_i, c_i)$ (component index, reliability improvement, cost), the optimal upgrade portfolio solves:

$$\max_{\mathcal{S} \subseteq [n]} \Delta R_s(\mathcal{S}) = R_s(\{p_i + \Delta R_i\}_{i \in \mathcal{S}}) - R_s(\mathbf{p}), \quad \text{s.t.} \quad \sum_{i \in \mathcal{S}} c_i \le B.$$

Under the *greedy approximation*: rank by $I_i^B \Delta R_i / c_i$ and select in order until budget is exhausted. This greedy policy is a $(1 - 1/e)$-approximation of the optimal (since ΔR_s is submodular in the selected set — the marginal benefit of adding component i decreases as more components are already upgraded).

Example — Importance-based maintenance prioritisation for a substation

A power substation has 8 components. Birnbaum importances at $t = 10$ years (computed from CTMC): $I_1^B = 0.42$, $I_3^B = 0.31$, $I_5^B = 0.28$, others < 0.10. Planned upgrades: component 1 ($\Delta R = 0.06$, cost \$40K), component 3 ($\Delta R = 0.04$, cost \$25K), component 5 ($\Delta R = 0.08$, cost \$50K). Budget: \$75K.
Greedy ranking by $I_i^B \Delta R_i / c_i$: component 3: $0.31 \times 0.04/25 = 4.96 \times 10^{-4}$; component 1: $0.42 \times 0.06/40 = 6.30 \times 10^{-4}$; component 5: $0.28 \times 0.08/50 = 4.48 \times 10^{-4}$.
Order: 1, 3, 5. Budget: select 1 (\$40K) + 3 (\$25K) = \$65K $\le$ \$75K. Component 5 is not affordable. System reliability improvement: $\Delta R_s \approx I_1^B \Delta R_1 + I_3^B \Delta R_3 = 0.42 \times 0.06 + 0.31 \times 0.04 = 0.025 + 0.012 = 0.037$ (3.7 percentage points).

Sensitivity analysis and importance in Monte Carlo. When system structure is too complex for analytical importance computation, Monte Carlo simulation estimates I_i^B by finite

differences: $\hat{I}_i^B = [\hat{R}_s(p_i + \delta) - \hat{R}_s(p_i - \delta)]/(2\delta)$. Each simulation run assigns component states by Bernoulli(p_j) draws. The variance of $\hat{I}_i^B$ from N simulation runs is $O(1/(N\delta^2))$ — variance increases for small perturbations, requiring a large N.

Global sensitivity analysis and Sobol' indices. For a system with reliability $R_s = f(p_1, \ldots, p_n)$ and uncertain component reliabilities p_i (random), the *first-order Sobol' index*:

$$S_i = \frac{\mathrm{Var}_{p_i}\left[\mathbb{E}_{p_{-i}}[R_s \mid p_i]\right]}{\mathrm{Var}[R_s]},$$

measures the fraction of system reliability variance attributable to p_i alone. The *total index*:

$$T_i = 1 - \frac{\mathrm{Var}_{p_{-i}}\left[\mathbb{E}_{p_i}[R_s \mid p_{-i}]\right]}{\mathrm{Var}[R_s]},$$

includes all interactions involving p_i. Under independence: $\sum_i S_i \leq 1 \leq \sum_i T_i$. The difference $T_i - S_i$ quantifies interaction effects (CCF, redundancy sharing).

Monte Carlo estimation of Sobol' indices. Saltelli's estimator uses two $N \times n$ sample matrices $\mathbf{A}$ and $\mathbf{B}$ (each row a sample from the p_i distribution):

$$\hat{S}_i = \frac{\frac{1}{N}\sum_{k=1}^{N} f(\mathbf{B}_k)[f(\mathbf{A}_k^{(i)}) - f(\mathbf{A}_k)]}{\hat{\mathrm{Var}}[R_s]},$$

where $\mathbf{A}_k^{(i)}$ has all columns from $\mathbf{A}$ except column i which comes from $\mathbf{B}$. Total indices use the alternate matrix $\mathbf{A}_k^{\sim i}$ (column i from $\mathbf{A}$, rest from $\mathbf{B}$). This requires $N(n+2)$ model evaluations — feasible for analytic reliability polynomials.

Example — Sobol' indices for a 5-component bridge system

Component reliabilities: $p_i = p + \epsilon_i$ with $p = 0.9$, $\epsilon_i \sim \mathrm{Uniform}(-0.05, 0.05)$ i.i.d. $N = 10000$ Sobol' samples. Results:

Component	S_i	T_i	$T_i - S_i$
1 (top-left)	0.18	0.22	0.04
2 (top-right)	0.18	0.22	0.04
3 (bot-left)	0.18	0.22	0.04
4 (bot-right)	0.18	0.22	0.04
5 (bridge)	0.09	0.14	0.05

$\sum S_i = 0.81 < 1$: 19% of variance is due to interactions. Component 5 (bridge) has the highest relative interaction effect ($T_5 - S_5 = 0.05$) because it connects both paths — its uncertainty interacts with all path components.

Importance in time-varying systems. For systems where component failure rates change over time (due to ageing or load variation), the Birnbaum importance $I_i^B(t)$ is time-varying. *Maintenance scheduling based on integrated importance*: repair component i at the time t_i^* that maximises $\int_{t_i^*}^{\infty} I_i^B(t)\Delta p_i(t)\, dt/c_i$, balancing the improvement in system reliability over the remaining life against maintenance cost.

8.4 Common-Cause Failures and Dependence Modelling

Common-Cause Failure (CCF)

A CCF event is a *dependent failure* in which two or more components fail simultaneously or within a short time window due to a shared root cause (same hardware design, common environment, shared maintenance error). CCF bypasses redundancy and is a dominant contributor to nuclear and aviation system failure probabilities.

α-factor model. For a group of n components, the α-factor model parameterises the fraction of failures involving exactly k components simultaneously:

$$\alpha_k = \frac{\text{fraction of failure events involving exactly } k \text{ components}}{\text{total failure events}}, \qquad \sum_{k=1}^{n} \alpha_k = 1.$$

Total failure rate for a single component: $\lambda_T = \lambda_{\text{observed total}}$. Rate of CCF events of size k: $\lambda_k = \frac{\binom{n}{k}^{-1} \alpha_k \lambda_T}{\sum_{j=1}^{n} \binom{n}{j}^{-1} \alpha_j}$.

MGL (Multiple Greek Letter) model. A hierarchical model with parameters $\beta, \gamma, \delta, \ldots$ representing conditional probabilities of CCF propagation to larger groups. For $n = 4$: $\lambda_1^{CCF} = (1-\beta)\lambda_T$; $\lambda_2^{CCF} = \beta(1-\gamma)\lambda_T/3$; $\lambda_3^{CCF} = \beta\gamma(1-\delta)\lambda_T/3$; $\lambda_4^{CCF} = \beta\gamma\delta\lambda_T/4$.

Copula models for statistical dependence. For stochastic dependence between component failure times $(T_1, \ldots, T_n)$, the joint distribution is specified via a copula C:

$$F(t_1, \ldots, t_n) = C(F_1(t_1), \ldots, F_n(t_n)).$$

Common choices: Gaussian copula (tail independence), Clayton copula ($C(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$, lower tail dependence), Gumbel copula (upper tail dependence). System failure time $T_{\min} = \min_i T_i$ has CDF computable from the copula and marginals.

Example — β-factor model for diesel generator CCF

A nuclear plant has 4 emergency diesel generators (EDG). Independent failure rate: $\lambda = 3 \times 10^{-3}$/demand. β-factor estimate from industry data: $\hat{\beta} = 0.08$.
Independent failure probability per demand: $q_I = (1-\beta)\lambda = 0.92 \times 3 \times 10^{-3} = 2.76 \times 10^{-3}$. CCF probability (all 4 fail): $q_C = \beta\lambda = 0.24 \times 10^{-3}$ (rough estimate for group of 4).
System of 4 in parallel: unreliability $\approx q_I^4 + q_C = (2.76 \times 10^{-3})^4 + 2.4 \times 10^{-4} \approx 5.8 \times 10^{-11} + 2.4 \times 10^{-4} \approx 2.4 \times 10^{-4}$.
Without CCF: $q_I^4 \approx 5.8 \times 10^{-11}$ — negligible. CCF dominates by 9 orders of magnitude!

Defence against CCF. The fundamental countermeasure is *diversity*: using components of different design, manufacturer, or technology to eliminate common failure mechanisms. Spatial separation addresses common environmental causes (fire, flooding). Independent testing and maintenance prevents correlated human errors.

Probability of failure on demand (PFD) for safety systems. For a safety instrumented system (SIS) that is tested periodically (interval τ), the average PFD is:

$$\text{PFD}_{\text{avg}} = \frac{1}{\tau} \int_0^{\tau} q(t)\, dt,$$

where $q(t) = F(t)$ (for a component not tested since $t = 0$). For exponential $q(t) = 1 - e^{-\lambda t} \approx \lambda t$ (for $\lambda\tau \ll 1$): $\text{PFD}_{\text{avg}} \approx \lambda\tau/2$.

IEC 61508 Safety Integrity Levels (SIL).

SIL	PFD range	RRF $= 1/\text{PFD}$
SIL 1	$[10^{-2}, 10^{-1})$	10–100
SIL 2	$[10^{-3}, 10^{-2})$	100–1000
SIL 3	$[10^{-4}, 10^{-3})$	1000–10000
SIL 4	$[10^{-5}, 10^{-4})$	10000–100000

For a 1oo2 (1-out-of-2, i.e. parallel) architecture: PFD $= (\lambda\tau/2)^2$ — one order of magnitude improvement per added redundant channel.

Systematic and random failures. IEC 61508 distinguishes:

- *Random hardware failures*: covered by the quantitative PFD analysis.
- *Systematic failures*: caused by design errors, specification errors, or common manufacturing defects. Not quantifiable by PFD; addressed by process requirements (V&V, independence of design teams, etc.).
- *CCF*: addressed by the β-factor adjustment (Section 8.4).

Markov Model for SIS with Detected/Undetected Failures

Two failure modes: *safe detected* (SD, rate λ_{SD}, revealed by online diagnostics) and *dangerous undetected* (DU, rate λ_{DU}, revealed only at proof test). CTMC states: {OK, SD-failure, DU-failure}. Steady-state PFD:

$$\text{PFD}_{\text{avg}} \approx \frac{\lambda_{\text{DU}}\tau}{2} + \frac{\lambda_{\text{DU}}}{\mu_{\text{repair}}}$$

(first term: time-averaged probability of being in DU state between tests; second term: probability of remaining in DU state during repair). The second term is often negligible for short repair times.

Example — SIL verification for a gas turbine trip system

1oo2 voting logic, $\lambda_{\text{DU}} = 8 \times 10^{-6}$/h per channel, $\tau = 8760$ h (annual proof test), $\beta = 0.05$ (CCF fraction).
Independent channel $\text{PFD}_1 = (1-\beta)\lambda_{\text{DU}}\tau/2 = 0.95 \times 8 \times 10^{-6} \times 8760/2 = 0.0333$.
1oo2 without CCF: $\text{PFD}_{1oo2} = \text{PFD}_1^2 = 0.00111$.
With CCF ($\beta = 0.05$): $\text{PFD}_{1oo2}^{\text{CCF}} = \text{PFD}_1^2 + \beta\lambda_{\text{DU}}\tau/2 = 0.00111 + 0.05 \times 8 \times 10^{-6} \times 8760/2 = 0.00111 + 0.00175 = 0.00286$.
Target SIL 2: PFD $\in [10^{-3}, 10^{-2})$. With CCF, the system achieves SIL 2 (PFD = 2.86×10^{-3}) but fails without considering CCF the computed value would be 1.1×10^{-3} — barely SIL 2. CCF is the dominant contributor (61% of PFD).

Network reliability and multi-state links. In network reliability problems (e.g. communication networks, power grids), edges have random capacities or failure states, and the system metric is *2-terminal reliability* (source-to-sink connectivity) or *all-terminal reliability* (all nodes connected). Exact computation is #P-hard; the standard approach is Monte Carlo simulation with importance sampling: bias the sampling toward failure states using the cross-entropy method to minimise variance.

Dependent failure modelling via Bayesian networks. A Bayesian network (BN) repre-

sents dependencies among component failures via a directed acyclic graph (DAG) with conditional probability tables (CPTs). For a system with n components, the joint failure probability:

$$\Pr(x_1, \ldots, x_n) = \prod_{i=1}^{n} \Pr(x_i \mid x_{\text{pa}(i)}),$$

where $\text{pa}(i)$ are the parents of node i in the DAG. System reliability is obtained by summing over all failure configurations: $R_s = \sum_{\mathbf{x}:\phi(\mathbf{x})=1} \Pr(\mathbf{x})$. The BN handles CCF naturally: the common cause event is a parent node whose failure probabilistically causes multiple child component failures.

Dynamic Bayesian network (DBN) for degradation. A time-sliced BN models the joint evolution of component degradation states $\mathbf{X}_t = (X_{1t}, \ldots, X_{nt})$:

$$\Pr(\mathbf{X}_t \mid \mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots) = \Pr(\mathbf{X}_t \mid \mathbf{X}_{t-1}) \quad \text{(Markov assumption)}.$$

The DBN unrolls over time, representing the entire degradation trajectory. Inference (filtering, smoothing, prediction) is performed via the junction tree algorithm. The DBN is particularly powerful for multi-component systems with shared degradation drivers (temperature, vibration, humidity) that influence multiple components simultaneously.

Copula-based CCF quantification. For two components with marginal failure times $T_1 \sim F_1$ and $T_2 \sim F_2$, their joint survival is $\Pr(T_1 > t_1, T_2 > t_2) = C(S_1(t_1), S_2(t_2))$ for a copula C. Common cause failure probability (both fail within interval Δt):

$$P_{\text{CCF}} = \Pr(T_1 \le \Delta t, T_2 \le \Delta t) = C(F_1(\Delta t), F_2(\Delta t)).$$

For independent components ($C = \Pi$): $P_{\text{CCF}} = F_1 F_2$. For upper Fréchet (perfect positive dependence): $P_{\text{CCF}} = \min(F_1, F_2)$. The Clayton copula with parameter θ interpolates: $C_\theta(u, v) = (\max(u^{-\theta} + v^{-\theta} - 1, 0))^{-1/\theta}$, with P_{CCF} increasing in θ (lower tail dependence).

Example — Copula-based CCF for pump seal pairs

Two pump seals have identical Weibull distributions ($k = 2$, $\mu = 5000\,\text{h}$). Maintenance interval: $\tau = 8760$ h (annual). $F(\tau) = 1 - e^{-(\tau/\mu\Gamma(1.5))^2} \approx 0.127$.
Independent ($C = \Pi$): $P_{\text{CCF}} = 0.127^2 = 0.016$. Clayton ($\theta = 2$, Kendall's $\tau = 0.5$): $C_2(0.127, 0.127) = (2 \times 0.127^{-2} - 1)^{-0.5} = (2 \times 62.1 - 1)^{-0.5} = 123.2^{-0.5} = 0.090$.
Fréchet upper ($C = M$): $\min(0.127, 0.127) = 0.127$.
With Clayton ($\theta = 2$): $P_{\text{CCF}} = 0.090$ vs. 0.016 (independence) — a 5.6× increase from dependence. The β factor implied: $\beta = P_{\text{CCF}}/F(\tau) = 0.090/0.127 = 0.71$ — extremely high compared to typical empirical $\beta \approx 0.05$–0.10. The copula model shows that strong tail dependence is more realistic than the constant-β model.

Environmental and load-sharing dependencies. Beyond purely statistical dependencies, physical coupling creates reliability interdependence:

- *Load-sharing*: surviving components carry higher load when one fails, increasing their failure rates. The surviving component failure rate: $\lambda_{\text{shared}} = \lambda_{\text{original}} \times (n/(n-k))^m$ for k failed components (m = stress exponent).
- *Warm environment coupling*: heat generated by component i raises the temperature of adjacent component j, accelerating j's Arrhenius-governed failure rate.

Multi-objective RAP: reliability vs. weight vs. volume. For aerospace and space systems,

the RAP typically involves three objectives simultaneously:

$$\max R_s(\mathbf{n}), \quad \min W(\mathbf{n}) = \sum_j n_j w_j, \quad \min V(\mathbf{n}) = \sum_j n_j v_j,$$

subject to $C(\mathbf{n}) \leq B$. The 3D Pareto frontier is computed by the ε-constraint method: fix $W \leq W_{\max}$ and $V \leq V_{\max}$, then solve the single-objective RAP; sweep $(W_{\max}, V_{\max})$ over a grid. Each Pareto point represents a different reliability-weight-volume tradeoff.

Surrogate-based optimisation for expensive RAP. When the reliability function $R_s(\mathbf{n})$ requires simulation (Monte Carlo) to evaluate — e.g. for non-series-parallel systems with complex failure dependencies — each evaluation is expensive. Surrogate models (Kriging, radial basis functions) approximate $R_s(\mathbf{n})$ from a small design of experiments (Latin Hypercube Sampling), then optimise the surrogate cheaply. Accuracy is assessed by cross-validation; new sample points are added where uncertainty is high (Expected Improvement criterion).

Genetic algorithm for large-scale RAP. For $k = 30$ stages and budget $B = 100$, the DP table has $30 \times 100 = 3000$ cells — still manageable. For $k = 100$ and non-series-parallel topology: standard DP breaks down. A genetic algorithm with:

- Chromosome: integer vector $(n_1, \ldots, n_k)$ with $n_j \geq 1$.
- Fitness: $R_s(\mathbf{n})$ if $C(\mathbf{n}) \leq B$, else $R_s(\mathbf{n}) - \lambda_{\text{pen}}(C(\mathbf{n}) - B)^2$ (penalty).
- Crossover: two-point on integer strings.
- Mutation: increment or decrement one n_j by 1, with probability $p_m = 1/k$.
- Selection: tournament ($k_{\text{tourn}} = 3$).
- Population: 200 individuals, 1000 generations.

Achieves within 1% of known optimal for $k \leq 50$ benchmark instances in < 10 seconds.

Example — Pareto frontier for satellite power system RAP

Satellite power system: $k = 4$ subsystems (solar array, battery, converter, distribution bus), $p_j = [0.92, 0.85, 0.90, 0.88]$, $c_j^{\text{cost}} = [4, 3, 2, 1]$, $w_j^{\text{weight}} = [3, 5, 2, 1]$ kg/unit. Budget: \$15M, weight limit: 40 kg.
Baseline ($\mathbf{n} = (1, 1, 1, 1)$): $R_s = 0.621$, $C = 10$, $W = 11$ kg.
Pareto frontier (sweep budget from 10 to 30 at weight limit 40 kg):

Budget	$\mathbf{n}^*$	R_s^*	C^*	W^* (kg)
10	$(1, 1, 1, 1)$	0.621	10	11
13	$(1, 1, 2, 3)$	0.760	13	16
16	$(2, 1, 2, 3)$	0.822	17	19
20	$(2, 2, 2, 5)$	0.882	21	31
25	$(3, 2, 2, 6)$	0.921	26	38

The *knee point* of the Pareto curve is at budget 16: from 13 to 16, reliability increases by 6.2 percentage points per 3M\$ (2.07 pp/M\$); from 16 to 20, the rate drops to 1.5 pp/M\$ (diminishing returns). The system engineer selects budget 16 as the cost-effective design point.

Universal generating function for multi-state RAP. For multi-state components (performance levels g_{jk}, $k = 0, \ldots, m_j$), the UGF of a parallel subsystem is computed by the max-

composition:

$$U_{\text{par}}(z) = \bigoplus_{j=1}^{n} U_j(z), \quad (p_a z^{g_a}) \oplus (p_b z^{g_b}) = p_a p_b z^{\max(g_a, g_b)}.$$

For a series subsystem: the min-composition $\otimes$. The system UGF $U_s(z) = \otimes_j U_j(z)$ gives the system performance distribution $\{P(G_s = g)\}$. System reliability at demand d: $R_s(d) = \sum_{g \geq d} P(G_s = g)$ — read directly from $U_s(z)$ by summing coefficients z^g for $g \geq d$.

8.5 Redundancy Allocation and Optimal Design

Redundancy Allocation Problem (RAP)

Given a k-stage series system with n_j parallel components at stage j, the RAP seeks to allocate redundancy $(n_1, \ldots, n_k)$ to maximise system reliability subject to weight, volume, or cost constraints:

$$\max_{n_1,\ldots,n_k} R_s(\mathbf{n}) = \prod_{j=1}^{k} R_j(n_j) \quad \text{s.t.} \quad \sum_{j=1}^{k} c_j n_j \leq C, \quad n_j \geq 1,\ n_j \in \mathbb{Z}^+,$$

where $R_j(n_j) = 1 - (1 - p_j)^{n_j}$ is stage j reliability with n_j components in parallel.

Lagrangian relaxation. Relax the constraint $\sum c_j n_j \leq C$ with multiplier $\mu \geq 0$:

$$L(\mathbf{n}, \mu) = -\sum_j \ln R_j(n_j) + \mu \left(\sum_j c_j n_j - C \right).$$

Solve the unconstrained problem over $n_j \in \mathbb{R}$ (continuous relaxation), then round to integers. Differentiate w.r.t. n_j:

$$-\frac{(1-p_j)^{n_j} \ln(1-p_j)}{1-(1-p_j)^{n_j}} + \mu c_j = 0.$$

This gives $n_j(\mu)$; adjust μ via bisection to satisfy the budget constraint.

Dynamic programming formulation.

$$f_j(W) = \max_{1 \leq n_j \leq \lfloor (W)/c_j \rfloor} R_j(n_j) \cdot f_{j+1}(W - c_j n_j),$$

with $f_{k+1}(W) = 1$ for all $W \geq 0$. The DP solves RAP exactly in $O(k \cdot C_{\max}/c_{\min})$ time.

Reliability–cost tradeoff curve. Plotting optimal R_s^* against budget C (the Pareto frontier) allows decision-makers to identify the *marginal cost of reliability*: dC^*/dR_s^* — the additional budget required for each additional unit of reliability. This marginal cost increases sharply as $R_s \to 1$ (diminishing returns of redundancy).

Example — Three-stage RAP solution

Series system, $k = 3$ stages, $c_j = [2, 3, 1]$ (units), budget $C = 10$, $p_j = [0.7, 0.8, 0.6]$. DP table excerpt:

$(1,1,1)$	6	0.700	0.800	0.600	0.336
$(2,1,1)$	8	0.910	0.800	0.600	0.437
$(1,1,3)$	9	0.700	0.800	0.936	0.524
$(2,1,2)$	9	0.910	0.800	0.840	0.612
$(1,2,2)$	9	0.700	0.960	0.840	0.565
$(2,1,2)$	9	0.910	0.800	0.840	0.612
$(1,1,4)$	10	0.700	0.800	0.9744	0.546
$(2,2,1)$	10	0.910	0.960	0.600	0.524
$(2,1,3)$	10	0.910	0.800	0.936	0.681

Optimal: $(n_1^*, n_2^*, n_3^*) = (2,1,3)$ with $R_s^* = 0.681$ at cost 10. Stage 3 (weakest, cheapest) receives most redundancy; stage 2 (strongest) receives only 1 component.

Mixed-integer programming (MIP) formulation. For large-scale RAP with non-series-parallel topologies, binary variables $y_{jk} = \mathbf{1}[n_j = k]$ give an exact MIP:

$$\max \ln R_s(\mathbf{n}), \quad \text{s.t.} \quad \sum_{k} k\, c_j\, y_{jk} \le C,\ \sum_{k} y_{jk} = 1,\ y_{jk} \in \{0,1\}.$$

State-of-the-art solvers (CPLEX, Gurobi) handle RAP instances with $k \le 50$ stages and $C \le 10^4$ in seconds, using LP relaxations and branch-and-bound.

Reliability–cost tradeoff: Non-dominated Pareto frontiers. In practice, reliability and cost are both objectives. Multi-objective optimisation of the RAP:

$$\min_{\mathbf{n}} [-R_s(\mathbf{n}),\ C(\mathbf{n})] \quad \text{s.t. } n_j \geq 1,\ n_j \in \mathbb{Z}^+.$$

The Pareto frontier of non-dominated solutions (no design is better in both objectives simultaneously) is computed by the ε-constraint method: fix $R_s \geq R^*$ and minimise C, then sweep R^* from $R_s^{\min}$ to $R_s^{\max}$. The knee point of the Pareto curve (maximum curvature) is a common engineering choice.

Genetic algorithms for RAP. For large k and non-series-parallel systems, evolutionary optimisation is practical:

1. *Encoding*: chromosome $= (n_1, n_2, \ldots, n_k)$.
2. *Fitness*: $f(\mathbf{n}) = R_s(\mathbf{n}) - \text{penalty}(\max(0, C(\mathbf{n}) - B))^2$.
3. *Operators*: crossover (two-point on integer strings), mutation (increment/decrement n_j by 1), selection (tournament).
4. *Convergence*: typically 500–2000 generations for $k \leq 100$.

GA solutions are heuristic but achieve within 1–2% of optimality in benchmark instances, much faster than exact DP for large k.

Standby redundancy. Active redundancy (all components running in parallel) vs. standby redundancy (backup is dormant until primary fails) differ in failure rates: a dormant component may have $\lambda_{\text{standby}} < \lambda_{\text{active}}$ (reduced stress). For a primary + one cold standby ($\lambda_{\text{standby}} = 0$):

$$S_{\text{cs}}(t) = e^{-\lambda t}(1 + \lambda t) \quad \text{(Erlang-2 distribution)}.$$

Compare with active parallel: $S_{\text{ap}}(t) = 2e^{-\lambda t} - e^{-2\lambda t}$. At $t = 1/\lambda$: $S_{\text{cs}} = 2/e \approx 0.736$ vs. $S_{\text{ap}} = 2/e - 1/e^2 \approx 0.601$. Cold standby is superior because the backup is not accumulating failures while idle.

Warm standby. For warm standby with $\lambda_{\text{sb}} \in (0, \lambda)$:

$$S_{\text{ws}}(t) = e^{-\lambda t} + \frac{\lambda}{\lambda - \lambda_{\text{sb}}}(e^{-\lambda_{\text{sb}} t} - e^{-\lambda t})\,.$$

As $\lambda_{\text{sb}} \to 0$: $S_{\text{ws}} \to S_{\text{cs}}$ (cold standby). As $\lambda_{\text{sb}} \to \lambda$: $S_{\text{ws}} \to S_{\text{ap}}$ (active parallel).

Example — Optimal redundancy allocation with standby option

Two-stage series system ($k = 2$), $p_j = 0.85$ per component, budget $C = 12$, $c_j = [2, 3]$. Each standby unit costs 50% more than an active unit (standby switch cost).
Active parallel option (n_j active): $R_j(n_j) = 1 - (0.15)^{n_j}$. Cold standby option ($m_j$ standbys with 1 active): $R_j^{\text{cs}}(m_j) = \sum_{k=0}^{m_j} \binom{m_j}{k}(1-p_j)^k p_j^{m_j-k} e^{-\lambda t}(1+\lambda t)^k/k!$... (simplified for static: $R_j^{\text{cs}} = 1 - (1-p_j)^{m_j+1}$, same as active parallel at fixed time for exponential — but better for IFR components).
For Weibull ($k = 2$) components, cold standby yields $\approx 12\%$ higher MTTF than active parallel at the same cost — a significant reliability gain for IFR failure modes.

Reliability allocation. Given a target system reliability R_s^* and k subsystems in series, the *equal apportionment* sets $R_j = (R_s^*)^{1/k}$. More sophisticated approaches:

AGREE method: weight subsystem reliability requirements by criticality w_j (fraction of total function provided) and complexity n_j (number of modules): $R_j = \exp(-w_j \ln R_s^* / n_j)$.

ARINC method: apportion failure rates proportional to current MTBF estimates: $\lambda_j^* = \lambda_j \cdot (\sum_i \lambda_i)/\lambda_s^{\max}$, where $\lambda_s^{\max}$ is the maximum allowable system failure rate.

Feasibility of allocation. A target R_s^* is achievable if there exists $\mathbf{n}$ with $C(\mathbf{n}) \le B$ and $R_s(\mathbf{n}) \ge R_s^*$. The maximum achievable reliability at budget B defines the *efficiency frontier* and is computed by the RAP with reversed objective.

Reliability growth testing and demonstration. After design changes, the system must demonstrate improved reliability. The AMSAA reliability growth planning model sets growth targets:

$$\lambda_{\text{goal}}(T_{\text{test}}) = \lambda_{\text{initial}} \cdot \left(\frac{T_{\text{test}}}{T_0} \right)^{-\alpha},$$

where α is the planned growth rate. The test duration T_{test} required to demonstrate $\lambda \le \lambda_{\text{goal}}$ at confidence $1-\gamma$ from a Crow–AMSAA model with n failures:

$$\chi^2_{2n+2,\gamma} / (2T_{\text{test}}) \le \lambda_{\text{goal}}.$$

Design for reliability (DfR) methodology. A systematic approach to embedding reliability into the design process:

1. *Reliability requirements allocation*: top-down apportionment to subsystems (Section 8.5).
2. *FMEA*: identify potential failure modes (Section 8.2 extension).
3. *Design validation tests*: HALT (Highly Accelerated Life Test), HASS (Highly Accelerated Stress Screening).
4. *Reliability prediction*: MIL-HDBK-217, Telcordia SR-332, or PoF models.
5. *Reliability demonstration test*: verify target before production.
6. *Field feedback*: infant mortality monitoring, warranty analysis, FMR (Field Mean Repair/Replacement Rate).

System of systems (SoS) reliability. In a SoS, independently operated component systems interact to deliver emergent capability. The SoS reliability model must account for: (i) *architectural connections* (interface protocols that can fail); (ii) *independent management* (each system has its own maintenance policy, potentially uncoordinated); (iii) *emergent behaviour* (the system-of-systems can degrade gracefully even if individual systems fail partially). Multi-agent modelling and agent-based simulation are common approaches for SoS reliability prediction.

Example — DfR applied to an automotive brake-by-wire system

Requirements: $R(10\,\text{yr}) \ge 0.9999$ (safety-critical). Allocated to 4 subsystems (ECU, actuator, sensor, power supply) in series. Per-subsystem target: $R_j(10\,\text{yr}) \ge 0.9999^{1/4} = 0.99998$, i.e. $\lambda_j \le 2 \times 10^{-9}$/h.
FMEA identifies 12 potential failure modes. HALT testing at elevated temperature ($80°C$) and vibration (20g) accelerates by AF $= 850$ (combined Arrhenius + Miner). Demonstration test: $n^* = \ln(0.05)/\ln(0.99998^{850}) \approx \ln(0.05)/(-850 \times 2 \times 10^{-9} \times 8760) \approx \ln(0.05)/(-1.49 \times 10^{-2}) = 201$ unit-hours. Run 5 units for 40 accelerated hours each with 0 failures: passes at 95% confidence.

Post-deployment: FMR tracking via OBD-II data. At 2 years, FMR indicates ECU failure mode not captured by FMEA (software watchdog timeout). Design change implemented (ECU firmware update). Reliability growth: λ_{ECU} reduces by factor of 3.

Reliability-centred maintenance (RCM). RCM (MSG-3 for civil aviation, MIL-STD-3034 for military) answers seven questions for each significant item:

1. What are the functions and performance standards?
2. How does it fail to fulfil these functions?
3. What causes each functional failure?
4. What happens when each failure occurs?
5. In what way does each failure matter?
6. What can be done to predict or prevent each failure?
7. What if no proactive task can be found?

The RCM decision diagram classifies each failure mode as: *on-condition* task (CBM), *scheduled restoration*, *scheduled discard*, or *run-to-failure* — based on the failure mode's safety consequences, economic consequences, and whether it has a well-defined P-F interval (detectable degradation window). The output is a *maintenance task list* that is both complete (covers all failure modes) and economically optimal.

Extended Case Study: Fleet Reliability Analysis for Helicopter Gearboxes

This case study illustrates the integrated application of survival analysis methods to a real-world fleet monitoring problem.

Setting. A helicopter operator monitors $n = 80$ main rotor gearboxes over a 12-year period. Each gearbox is installed at time zero and removed at failure or at the scheduled overhaul interval (hard-time limit of $T_{\text{HL}} = 3000$ flight hours, FH). Over the study period, 31 gearboxes failed before the hard-time limit (uncensored failures), and 49 reached the hard-time limit and were overhauled (right-censored at 3000 FH).

Step 1: Kaplan–Meier estimate. The KM survival function is computed from the 80 pairs (t_i, δ_i). The curve declines gradually from 1.0 at $t = 0$ to approximately 0.60 at the hard-time limit, reflecting that about 40% of gearboxes are expected to fail before 3000 FH.

Step 2: Parametric modelling. AIC comparison (Weibull: 312.4; Log-Normal: 315.1; Gamma: 314.0) favours the Weibull. MLE: $\hat{k} = 2.84$ (IFR — consistent with mechanical wear-out), $\hat{\lambda}^{-1} = 3920$ FH. The Weibull probability plot is linear ($p = 0.42$ in the Anderson–Darling test).

Step 3: Covariate analysis via Cox PH. Two covariates: x_1 = operating temperature index (OTI, standardised), x_2 = vibration severity index (VSI, standardised) from health and usage monitoring systems (HUMS). Cox PH MLE:

$$\hat{\boldsymbol{\beta}} = (0.41,\, 0.58), \quad \text{SE} = (0.17,\, 0.21), \quad \text{Hazard ratios}: e^{0.41} = 1.51,\ e^{0.58} = 1.79.$$

Interpretation: each SD increase in OTI raises the failure hazard by 51%; each SD increase in VSI raises it by 79%. Both covariates are significant at the 5% level. Schoenfeld residuals show no temporal trend (PH assumption holds, $p = 0.34$).

Step 4: Individual risk stratification. For a specific gearbox with $(\text{OTI} = 1.2, \text{VSI} = 0.8)$ (above average on both), the personalised survival function:

$$\hat{S}(t|\mathbf{x}) = [\hat{S}_0(t)]^{\exp(0.41\times 1.2 + 0.58\times 0.8)} = [\hat{S}_0(t)]^{\exp(0.956)} = [\hat{S}_0(t)]^{2.60}.$$

At $t = 2000$ FH: $\hat{S}_0(2000) = e^{-(0.000256\times 2000)^{2.84}} = e^{-0.189} = 0.828$. Personalised: $0.828^{2.60} = 0.574$. This gearbox has only a 57.4% probability of surviving to 2000 FH — below the fleet average of 82.8%.

Step 5: Maintenance decision. The operator uses the 5th-percentile survival time (B5 life) as the trigger for inspection: for the personalised model, $\hat{t}_{B5} = 1380$ FH (vs. fleet B5 of 1650 FH). The gearbox is flagged for an intermediate inspection at 1380 FH rather than waiting for the standard 1800 FH inspection interval.

Step 6: Prospective validation. Over the next 2 years, gearboxes in the top quartile of personalised risk (predicted $\hat{S}(2000|\mathbf{x}) < 0.65$) experience 3.2× more failures before 2000 FH compared to the bottom quartile. The Cox PH model's C-statistic (concordance) is 0.74 — good discrimination.

Lessons learned. (1) The Weibull shape $\hat{k} = 2.84 > 1$ confirms that preventive maintenance is worthwhile (IFR). (2) HUMS data (VSI) is the stronger predictor ($\hat{\beta}_2 = 0.58$) than temperature. (3) Personalised survival estimates enable risk-based inspection intervals rather than fixed hard-time limits, reducing both over-maintenance (for healthy gearboxes) and under-maintenance (for stressed ones).

Extended Case Study: System Reliability for a Wind Farm Power Converter

This case study integrates structure functions, importance measures, and redundancy allocation for a wind turbine power converter (WTPC) system.

System architecture. A WTPC consists of: generator-side converter (GSC), DC link capacitor bank (DCL), grid-side converter (GSI), filter inductor (FIL), and control unit (CTL). The logical structure:

- GSC and GSI are both required (series path 1: GSC–DCL–GSI).
- A bypass contactor (BYP) allows direct grid connection in emergency, bypassing the converter.
- System works if: (full conversion path) OR (bypass path is active and grid conditions permit).

Simplified: series system of GSC, DCL, GSI, FIL with a parallel bypass path through BYP.

Component failure rates. Based on IEC 61400-26 and field data from 500-turbine fleet:

Component	λ_i (failures/year)	p_i (1-year)
GSC	0.12	0.887
DCL	0.03	0.970
GSI	0.11	0.896
FIL	0.02	0.980
BYP	0.04	0.961
CTL	0.08	0.923

(The control unit CTL is series with everything — a failure stops the system regardless of other components.)

Structure function. $\phi(\mathbf{x}) = x_{\text{CTL}} \cdot [\underbrace{x_{\text{GSC}} \cdot x_{\text{DCL}} \cdot x_{\text{GSI}} \cdot x_{\text{FIL}}}_{\text{main path}} + (1 - \underbrace{x_{\text{GSC}} \cdot x_{\text{DCL}} \cdot x_{\text{GSI}} \cdot x_{\text{FIL}}}) \cdot x_{\text{BYP}}]$

Simplifying: system works if CTL works AND (main path works OR bypass works given main fails).

Exact system reliability.

$$R_s = p_{\text{CTL}}[p_M + (1 - p_M)p_{\text{BYP}}],$$

where $p_M = p_{\text{GSC}} \cdot p_{\text{DCL}} \cdot p_{\text{GSI}} \cdot p_{\text{FIL}} = 0.887 \times 0.970 \times 0.896 \times 0.980 = 0.754$.

$R_s = 0.923 \times [0.754 + 0.246 \times 0.961] = 0.923 \times [0.754 + 0.236] = 0.923 \times 0.990 = 0.914$.

Importance measures. $I^B_{\text{CTL}} = R_s(1_{\text{CTL}}) - R_s(0_{\text{CTL}}) = 0.990 - 0 = 0.990$ (CTL is critical for *every* system state — it is the single dominant component).

$I^B_{\text{GSC}} = R_s(1_{\text{GSC}}) - R_s(0_{\text{GSC}})$. $R_s(1_{\text{GSC}}) = 0.923 \times [0.887_{\to 1} \times 0.970 \times 0.896 \times 0.980 + (1 - \ldots) \times 0.961] = 0.923 \times 0.990 = 0.914$ wait — with GSC=1 always: $p^+_M = 1 \times 0.970 \times 0.896 \times 0.980 = 0.850$; $R_s(1_{GSC}) = 0.923 \times [0.850 + 0.150 \times 0.961] = 0.923 \times 0.994 = 0.917$. With GSC=0: $p^-_M = 0$; $R_s(0_{GSC}) = 0.923 \times [0 + 1 \times 0.961] = 0.923 \times 0.961 = 0.887$. $I^B_{\text{GSC}} = 0.917 - 0.887 = 0.030$.

FV importance for CTL: CTL appears in all MCS (it is a singleton cut set): $I^{FV}_{\text{CTL}} = (1 - p_{\text{CTL}})/(1 - R_s) = 0.077/0.086 = 0.895$ — 89.5% of system unreliability is attributable to CTL failure alone.

Optimal redundancy allocation. Budget: 2 additional components (cost $c = 1$ per component). Greedy by $I_i^B \Delta R_i / c_i$:

- Add CTL redundancy (1oo2 CTL): $R_{\text{CTL}}^{1oo2} = 1 - (1 - 0.923)^2 = 0.994$. $\Delta R_s \approx I_{\text{CTL}}^B \times \Delta p_{\text{CTL}} = 0.990 \times (0.994 - 0.923) = 0.070$.
- Add GSC redundancy: $\Delta R_s = 0.030 \times 0.028 = 0.00084$.

Optimal first choice: add CTL redundancy (by far the dominant component). After CTL is made 1oo2: new $R_s = 0.994 \times 0.990 = 0.984$. Second component: add GSI redundancy ($I_{\text{GSI}}^B \approx 0.029$): $R_s \to 0.986$. Final system reliability: 0.986 (from 0.914) — a 72-point improvement in availability.

Annual energy loss reduction. At €100/MWh and 8760 h/year, rated power 3MW:

- Before: unavailability $= 1 - 0.914 = 0.086$. Expected energy not generated: $0.086 \times 3 \times 8760 = 2258$ MWh/year × €100 = €225,800/year.
- After redundancy: unavailability $= 0.014$. Expected energy not generated: 367 MWh/year × €100 = €36,700/year.
- Annual saving: €189,100/year. Payback on 2 redundant components (€20,000 installed): < 2 months.

Bayesian Reliability Analysis

Bayesian methods provide a principled framework for combining prior engineering knowledge with observed failure data, particularly valuable when data are sparse (few failures, short test durations).

Bayesian Reliability Framework

For a parametric failure-time model with parameters $\boldsymbol{\theta}$, the Bayesian update is:

$$p(\boldsymbol{\theta} \mid \text{data}) \propto L(\text{data} \mid \boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta}),$$

where L is the likelihood (incorporating censoring), $\pi(\boldsymbol{\theta})$ is the prior (encoding pre-test engineering knowledge), and $p(\boldsymbol{\theta}|\text{data})$ is the posterior. The *posterior predictive* reliability at time t:

$$\hat{R}(t) = \int S(t \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \text{data})\, d\boldsymbol{\theta},$$

averages over parameter uncertainty — unlike the MLE plug-in estimate $S(t \mid \hat{\boldsymbol{\theta}})$ which ignores it.

Conjugate priors for reliability models.

- *Exponential likelihood, Gamma prior*: $T \sim \text{Exp}(\lambda)$, $\lambda \sim \text{Gamma}(a, b)$. Posterior: $\lambda|\text{data} \sim \text{Gamma}(a + n_f, b + \sum t_i)$, where n_f failures and $\sum t_i$ total test time (exact closed form).
- *Binomial reliability, Beta prior*: n tests, k failures, $p = 1 - R \sim \text{Beta}(\alpha, \beta)$. Posterior: $p|\text{data} \sim \text{Beta}(\alpha + k, \beta + n - k)$. Posterior mean reliability: $\hat{R} = (\beta + n - k)/(\alpha + \beta + n)$.
- *Weibull likelihood*: no conjugate prior. Use MCMC (Metropolis–Hastings or HMC/NUTS) with weakly informative priors $k \sim \text{Gamma}(2, 0.5)$ (mean 4, concentrating IFR zone) and $\ln \lambda \sim \mathcal{N}(0, 2)$.

Prior elicitation from engineering knowledge. Several structured methods exist:

1. *Quantile method*: ask the expert "What is the 10th-percentile life?" and "What is the 90th-percentile life?" Fit a log-normal prior to match: $\mu_{\ln T} = (\ln t_{90} + \ln t_{10})/2$, $\sigma_{\ln T} = (\ln t_{90} - \ln t_{10})/3.29$.
2. *Predictive elicitation*: ask about observable quantities ("How many out of 10 components will fail before 1000 hours?") rather than abstract parameters. The expert's distribution over counts maps to a prior on $p = F(1000)$.
3. *Historical data*: use MLE from a previous product generation as the prior mean; inflate variance by a factor reflecting the design change magnitude.

Bayesian prediction intervals for future failures. Given posterior $p(\boldsymbol{\theta}|\text{data})$, the posterior predictive probability that a new component fails before time t:

$$\Pr(T_{\text{new}} \le t|\text{data}) = \int F(t|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\text{data})\, d\boldsymbol{\theta}.$$

The $(1 - \alpha)$ highest posterior density (HPD) interval for T_{new} is the shortest interval containing probability $1 - \alpha$ under the predictive distribution. For the Exponential-Gamma conjugate model: $T_{\text{new}}|\text{data} \sim \text{Lomax}(a + n_f, b + \sum t_i)$ (Pareto type II), with closed-form CDF.

Example — Bayesian reliability update for spacecraft thruster

Prior: exponential lifetime with $\lambda \sim \text{Gamma}(a = 2, b = 1000)$ (mean $\mathbb{E}[\lambda] = 0.002$/h, based on similar thruster design). Test data: 5 thrusters tested for 2000 h each; 1 failure at 1400 h (4 censored at 2000 h).
Posterior: $\lambda|\text{data} \sim \text{Gamma}(a + n_f, b + \sum t_i) = \text{Gamma}(3, 1000 + 4 \times 2000 + 1400) = \text{Gamma}(3, 9400)$.
Posterior mean: $\hat{\lambda} = 3/9400 = 3.19 \times 10^{-4}$/h (compared to prior mean 2×10^{-3}/h — test data strongly pulls down the failure rate estimate).
Posterior predictive reliability at 5000 h: $\hat{R}(5000) = \int_0^\infty e^{-5000\lambda} \text{Gamma}(\lambda; 3, 9400)\, d\lambda = \left(\frac{9400}{9400+5000}\right)^3 = \left(\frac{9400}{14400}\right)^3 = 0.653^3 = 0.279$.
Compare MLE plug-in: $\hat{R}_{\text{MLE}}(5000) = e^{-5000\hat{\lambda}_{\text{MLE}}} = e^{-5000/9400} \approx 0.589$. The Bayesian estimate is substantially more conservative because it accounts for parameter uncertainty — with only 1 observed failure, the true λ could be much higher than the MLE suggests.
95% HPD credible interval for $R(5000)$: $[0.08, 0.56]$ — very wide, reflecting the limited data.

Sequential Bayesian updating. As new failure data arrives, the posterior from the previous analysis becomes the new prior:

$$\pi_{k+1}(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\text{data}_1, \ldots, \text{data}_k) \propto L(\text{data}_{k+1}|\boldsymbol{\theta})\, \pi_k(\boldsymbol{\theta}).$$

For the exponential-Gamma conjugate: $(a_k, b_k) \to (a_k + \Delta n_f, b_k + \Delta T_{\text{test}})$ at each update — an $O(1)$ operation. This enables real-time reliability tracking as fleet data accumulates, converging the posterior toward the true λ at rate $O(1/\sqrt{n})$.

Empirical Bayes. When multiple components share a common prior but have individual failure histories, the hyperparameters (a, b) of the prior are estimated from the aggregate data (marginal likelihood maximisation). This "borrows strength" across components: well-tested components with many failures have a strong local posterior; components with few failures are regularised toward the group estimate. The resulting estimators have smaller MSE than individual MLEs — a James–Stein-type shrinkage result for reliability estimation.

Accelerated Life Testing and Reliability Extrapolation

Accelerated life testing (ALT) applies elevated stress levels to components, inducing failures in a fraction of the time required at nominal operating conditions, then extrapolates to estimate reliability under normal use.

Acceleration Models

An acceleration model relates the failure-time distribution parameters to the applied stress s:

- **Arrhenius**: $\ln \mu(T) = \beta_0 + \beta_1/T$ (temperature T in Kelvin). $\mu(T)$ = mean failure time. Activation energy: $E_a = k_B \beta_1$ where $k_B = 8.617 \times 10^{-5}$ eV/K.
- **Inverse Power Law**: $\mu(V) = \gamma V^{-p}$ (voltage V, power $p > 0$). Common for dielectric breakdown.
- **Eyring**: $\mu(T, V) = A/T \cdot e^{B/T} \cdot e^{(C+D/T)V}$. Combines temperature and voltage.
- **Coffin–Manson** (thermal cycling): $N_f = A(\Delta T)^{-p}$ where ΔT is temperature swing.

The *acceleration factor* $AF = \mu_{\text{normal}}/\mu_{\text{test}}$ converts test life to field life.

Accelerated Failure Time formulation. Any acceleration model can be cast as an AFT model:

$$\ln T = \mu(s) + \sigma\epsilon,$$

where $\mu(s)$ is the log-mean at stress s (given by the acceleration model) and ϵ has a fixed distribution (Smallest Extreme Value for Weibull AFT; Normal for Log-Normal AFT). The Weibull AFT with Arrhenius:

$$\ln T = \beta_0 + \frac{\beta_1}{T_{\text{K}}} + \sigma\epsilon, \quad T_{\text{failure}} = \exp(\beta_0 + \beta_1/T_K) \cdot U^{1/\hat{k}},$$

where $U \sim \text{Exp}(1)$ and $\hat{k} = 1/\sigma$. The shape parameter k is assumed constant across stress levels (stress only affects the scale — a testable assumption via parallel probability plots at each stress level).

MLE for ALT with multiple stress levels. Data: n_j units tested at stress s_j ($j = 1, \ldots, J$), yielding failure times t_{ij} (possibly censored). Log-likelihood:

$$\ell(\boldsymbol{\beta}, k) = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \left[\delta_{ij} \ln f(t_{ij}; \mu(s_j), k) + (1 - \delta_{ij}) \ln S(t_{ij}; \mu(s_j), k)\right].$$

For Weibull AFT with Arrhenius, $\boldsymbol{\beta} = (\beta_0, \beta_1)$ and k: 3 parameters estimated from potentially hundreds of observations.

Confidence intervals via profile likelihood. At the use stress s_0, the p-th quantile life $t_p(s_0) = \exp(\mu(s_0)) \cdot [-\ln(1-p)]^{1/k}$. The profile likelihood CI for $t_p(s_0)$ requires:

1. Compute $\ell_{\max}(\boldsymbol{\beta}, k)$ (global MLE).
2. For each candidate t^*: reparametrise and maximise the constrained likelihood over $(\boldsymbol{\beta}, k)$ subject to $t_p(s_0; \boldsymbol{\beta}, k) = t^*$.
3. The $(1-\alpha)$ CI: $\{t^* : 2[\ell_{\max} - \ell_{\text{profile}}(t^*)] \leq \chi^2_{1,\alpha}\}$.

Example — Arrhenius ALT for LED luminaires

12V LED luminaires tested at 3 temperatures: 70°C (14 units, 5 failures in 2000h), 85°C (12 units, 9 failures in 1500h), 100°C (10 units, 10 failures in 1000h). Failure criterion: $> 50\%$ lumen depreciation.
Weibull AFT with Arrhenius MLE: $\hat{\beta}_0 = -14.2$, $\hat{\beta}_1 = 8120$ K ($\hat{E}_a = 8120 \times 8.617 \times 10^{-5} = 0.70$ eV — reasonable for LED degradation), $\hat{k} = 2.1$ (IFR, consistent with wear-out).
Acceleration factor from 85°C to 25°C (nominal use): $AF = \exp[8120\,(\frac{1}{298} - \frac{1}{358})] = \exp[8120 \times 5.64 \times 10^{-4}] = e^{4.58} = 97.5$.
Predicted B10 life at 25°C: at 85°C, $\hat{t}_{10}(85°\text{C}) = \exp(-14.2 + 8120/358) \cdot [-\ln 0.9]^{1/2.1} = \exp(8.50) \times 0.318 = 4900 \times 0.318 = 1558$ h. $t_{10}(25°\text{C}) = 1558 \times 97.5 = 151{,}905$ h ≈ 17.3 years.
95% profile likelihood CI for $t_{10}(25°\text{C})$: $[9.8, 28.1]$ years — wide uncertainty reflecting the large extrapolation distance.

Step-stress testing. Instead of separate groups at constant stresses, step-stress ALT increases stress at pre-specified times for each unit: stress s_1 for $[0, \tau_1)$, then s_2 for $[\tau_1, \tau_2)$, etc. The cumulative exposure model (Nelson, 1980) assumes that at each stress change, the component's residual life under the new stress equals what it would be if it had aged under the new stress from $t = 0$ but with an initial "effective age" matching the cumulative damage. Efficient step-stress designs (optimal τ_1, τ_2, n_1, n_2 allocations) are available from D-optimal design theory.

Degradation-based ALT. Instead of waiting for failure, measure the degradation path $X(t)$ at each stress level and extrapolate the first-passage time. Under the Wiener process model: the drift $\mu(s) \propto \exp(\beta_1 / T_K)$ (Arrhenius). Fitting the degradation paths at multiple temperatures gives $\hat{\beta}_1$ and extrapolates to $\mu(s_0)$. This requires fewer units and shorter test durations than failure-based ALT, at the cost of needing a valid degradation-to-failure model.

Cross-Volume Connections and Methodological Synthesis

Connection to Volume I: Extreme Values and Tail Risk

The extreme value theory of Volume I and the survival analysis of this volume are deeply connected through the mathematics of distributional tails.

Minimum vs. maximum extremes. Volume I focused on the maximum of i.i.d. observations $M_n = \max(X_1, \ldots, X_n)$, governed by the GEV distribution. System reliability is determined by the *minimum*: $T_{\min} = \min(T_1, \ldots, T_n)$ (series system failure time). By the min-stability property, if $T_i \sim \text{Weibull}(\lambda, k)$, then $T_{\min} \sim \text{Weibull}(n^{1/k}\lambda, k)$. This is the exact counterpart of the max-stability of the Gumbel/Fréchet/Weibull GEV families: the Weibull distribution is both min-stable and has connections to the GEV through the transformation $T \to 1/T$.

Regular variation and the tail of the failure time distribution. The survival function $S(t)$ is *regularly varying* at ∞ with index $-\alpha$ if $S(t) = t^{-\alpha} L(t)$ for a slowly varying L. This is the failure-time analog of the Pareto tail. Regularly varying failure times arise in:

- Software reliability (power-law bug discovery rates).
- Social infrastructure failures (power grids, financial systems — Vol. IV material).
- Biological mortality at extreme ages (the Gompertz tail weakens, approaching a plateau — the "mortality deceleration" or Kannisto effect).

For regularly varying $S(t)$, the mean residual life $e(t) \sim t/(\alpha-1)$ for $\alpha > 1$ — increasing MRL (DFR behaviour), consistent with the aging-slowdown observed in centenarians.

Pickands–Balkema–de Haan theorem for exceedances. For a random variable T with right-heavy tail, the distribution of excesses $T-u \,|\, T > u$ converges (as $u \to \infty$) to the Generalised Pareto Distribution (GPD): $G_{\xi,\sigma}(x) = 1 - (1 + \xi x/\sigma)^{-1/\xi}$. Applied to failure times: the *conditional failure time given survival to time* u is approximately GPD for large u — a model-free result that does not assume parametric hazard shapes.

Applications to very long-lived components. Nuclear reactor pressure vessels are designed for 40–60 year lives. Estimating $S(40\,\text{yr})$ from laboratory data with test durations of ≤ 5 years requires extreme value extrapolation. The GPD tail fit provides:

$$\hat{S}(t) = \hat{S}(u)\left[1 + \hat{\xi}\frac{t-u}{\hat{\sigma}}\right]^{-1/\hat{\xi}}, \quad t > u,$$

with uncertainty quantified by profile likelihood confidence intervals on ξ and σ.

Connection to Volume III: Robust Optimisation and Uncertainty

Volume III addresses decisions under uncertainty — a setting directly motivated by the parameter estimation uncertainty quantified in this volume.

Distributional robustness in maintenance. When the failure time distribution F is uncertain (estimated from limited data), the optimal maintenance policy $\pi^*(F)$ may perform poorly for alternative plausible F. The *robust maintenance policy* solves:

$$\pi^{\text{robust}} = \arg\min_{\pi} \max_{F \in \mathcal{F}} g(\pi, F),$$

where $\mathcal{F}$ is an ambiguity set (e.g. all Weibull distributions with $k \in [\hat{k}-\delta, \hat{k}+\delta]$ at the 95% confidence level). For IFR distributions, the worst-case Weibull has the smallest k consistent with the data (slowest increasing hazard — most difficult to benefit from PM).

Chance-constrained maintenance. A reliability constraint $\Pr(S_{\text{system}}(T_{\text{mission}}) \geq r^*) \geq 1-\epsilon$ under uncertain component reliabilities $\mathbf{p}$ is a joint chance constraint:

$$\Pr(\phi(\{X_i \sim \text{Bernoulli}(p_i)\}) \geq 1 \text{ throughout } [0, T_m]) \geq 1-\epsilon.$$

Scenario approximation (Calafiore–Campi): draw $N = O(\log(1/\delta)/\epsilon)$ scenarios for the uncertain $\mathbf{p}$ and require the constraint to hold for all scenarios — gives a robust solution with high probability.

Connection to Volume IV: Cascading Failures and Network Resilience

The component-level failure models of this volume aggregate to network-level failure phenomena studied in Volume IV.

From independent components to cascading failures. When components share loads (Section 8.4: load-sharing CCF), the failure of one component increases the stress on others, potentially triggering a cascade. The avalanche model: each failed component redistributes its load w_i equally among its k intact neighbours. A neighbour j fails if its new load $w_j + w_i/k > w_j^{\text{thresh}}$. This is a threshold cascade on a network — studied in detail in Volume IV using mean-field approximations and percolation theory.

Network reliability and connectivity. For a network of n nodes with random link failures ($p_{\text{link}} = 1-q$), the system reliability (2-terminal or all-terminal) undergoes a phase transition

at a critical p_c^*: for $q > q_c^*$, the giant connected component disappears and reliability collapses. Near the threshold: $R_s \sim |q - q_c^*|^\beta$ (critical exponent β depending on network topology). For Erdős–Rényi graphs: $q_c^* = 1/\bar{d}$ where $\bar{d}$ is the mean degree.

Connection to Volume V: Rare Events and Large Deviations

The tail probabilities of failure time distributions connect directly to the large deviations theory of Volume V.

Large deviations for cumulative damage. Miner's damage $D_n = \sum_{i=1}^{n} d_i$ (sum of i.i.d. damage increments) satisfies the LDP:

$$\Pr(D_n/n \geq \delta) \approx e^{-nI(\delta)}, \quad I(\delta) = \sup_{\lambda \geq 0}[\lambda\delta - \ln \mathbb{E}[e^{\lambda d_i}]].$$

For small $\delta < \mathbb{E}[d_i]$: $I(\delta) > 0$ (exponentially unlikely to fail before expected Miner life). The Cramér rate function $I(\delta)$ characterises the tail of the damage distribution and gives precise asymptotic failure probabilities for large n.

Importance sampling from large deviations. The optimal importance sampling distribution for estimating $\Pr(D_n \geq n\ell)$ is the exponentially tilted distribution $\tilde{\Pr}(d) \propto e^{\lambda^* d} \Pr(d)$ where $\lambda^* = I'(\ell)^{-1}$ (the Legendre transform at ℓ). Under $\tilde{\Pr}$, $\mathbb{E}_{\tilde{\Pr}}[d] = \ell$ — so typical paths under the IS distribution reach the failure threshold, achieving exponential variance reduction compared to crude Monte Carlo.

Numerical Tour: Part 5 — From Raw Data to Survival Curve

The following worked tour walks through a complete survival analysis pipeline starting from raw (censored) failure data to a fully interpreted Cox model, using explicit arithmetic at every step.

Dataset. Twelve industrial pumps are observed over 400 operating days. Their (time, event) pairs are:

Unit	t_i (days)	δ_i	Unit	t_i (days)	δ_i
1	45	1	7	220	0
2	78	1	8	255	1
3	110	0	9	290	1
4	140	1	10	320	0
5	165	1	11	360	1
6	195	0	12	395	1

($\delta_i = 1$: failed; $\delta_i = 0$: censored at t_i.)

Step 1: Kaplan–Meier estimate. Ordered failure times: $45, 78, 140, 165, 255, 290, 360, 395$. At each failure:

t_j	n_j (at risk)	d_j (failures)	$1 - d_j/n_j$	$\hat{S}(t_j)$
45	12	1	11/12	0.9167
78	11	1	10/11	$0.9167 \times 10/11 = 0.8333$
140	9	1	8/9	$0.8333 \times 8/9 = 0.7407$
165	8	1	7/8	$0.7407 \times 7/8 = 0.6481$
255	6	1	5/6	$0.6481 \times 5/6 = 0.5401$
290	5	1	4/5	$0.5401 \times 4/5 = 0.4321$
360	3	1	2/3	$0.4321 \times 2/3 = 0.2881$
395	2	1	1/2	$0.2881 \times 1/2 = 0.1440$

Note: units censored at $t = 110, 195, 220, 320$ are removed from the risk set at their censoring time.

Step 2: Nelson–Aalen cumulative hazard.

$$\hat{H}(t_j) = \sum_{t_k \le t_j} d_k/n_k: \quad \hat{H}(45) = 1/12,\ \hat{H}(78) = 1/12{+}1/11 = 0.174,\ \ldots,\ \hat{H}(395) = \sum_j d_j/n_j = 1.989.$$

Compare with KM: $-\ln \hat{S}(395) = -\ln(0.144) = 1.937$ (close, as expected; NA slightly larger).

Step 3: Weibull parametric MLE. With the 8 exact failure times and 4 censored:

$$\ell(k, \lambda) = 8 \ln k + 8k \ln \lambda + (k-1) \sum_{i:\delta_i=1} \ln t_i - \lambda^k \sum_{i=1}^{12} t_i^k.$$

(The censored contribution enters only in $-\lambda^k t_i^k$, not in $\ln f(t_i)$.)

Numerical optimisation yields: $\hat{k} = 1.21$, $\hat{\lambda} = 0.00282$ day^{-1}. $\hat{\mu} = \Gamma(1 + 1/1.21)/\hat{\lambda} = \Gamma(1.827)/0.00282 = 0.9198/0.00282 = 326$ days.

Step 4: AIC comparison.

$$\ell_{\text{Weibull}} = -29.4, \quad \ell_{\text{Exponential}} = -31.2, \quad \text{AIC}_\text{W} = 62.8 < \text{AIC}_\text{E} = 64.4.$$

Weibull is preferred; $\hat{k} = 1.21 > 1$ indicates slight IFR (preventive maintenance marginally beneficial).

Step 5: Confidence intervals. Greenwood's formula at $t = 200$ days:

$$\hat{S}(200) = 0.432 \text{ (between } t_6 = 165 \text{ and } t_7 = 255), \quad \hat{V} = 0.432^2 \times \sum_{t_j \le 200} \frac{d_j}{n_j(n_j - d_j)}.$$

$\sum = \frac{1}{12\times11} + \frac{1}{11\times10} + \frac{1}{9\times8} + \frac{1}{8\times7} = 0.00758 + 0.00909 + 0.01389 + 0.01786 = 0.0484$. $\hat{V} = 0.432^2 \times 0.0484 = 0.00903$, $\hat{\sigma} = 0.095$. Log-log 95% CI: $\hat{S}(200)^{\exp(\pm 1.96 \times 0.095/[0.432 \times \ln 0.432])} = 0.432^{\exp(\pm 1.96 \times 0.095/0.352)} = 0.432^{\exp(\pm 0.529)} = [0.432^{1.697},\ 0.432^{0.589}] = [0.228,\ 0.640]$.

Step 6: Log-rank test vs. an independent control group. Suppose a control group of 12 identical pumps with no maintenance upgrade has KM curve $\hat{S}_{\text{control}}(t)$ with 10 failures (vs. 8 above). Log-rank $\chi^2 = 3.84$, $p = 0.050$ (marginal significance). Prentice–Wilcoxon: $\chi^2 = 4.21$, $p = 0.040$. The weighted test gives stronger evidence, consistent with the hazard ratio being larger early in the follow-up period — an early treatment benefit.

Formula Reference: Part 5 — Survival Analysis

Quantity	Formula	Notes
Hazard function	$h(t) = f(t)/S(t) = -d \ln S(t)/dt$	Definition
Cumulative hazard	$H(t) = -\ln S(t) = \int_0^t h(u)\,du$	Bijection
Survival from hazard	$S(t) = \exp(-H(t))$	Key link
Mean residual life	$e(t) = \int_t^\infty S(u)\,du / S(t)$	
MRL determines S	$S(t) = \frac{e(0)}{e(t)} \exp(-\int_0^t du/e(u))$	Inversion
Exponential	$h = \lambda$, $S = e^{-\lambda t}$, $\mu = 1/\lambda$	Memoryless
Weibull hazard	$h(t) = k\lambda(\lambda t)^{k-1}$	IFR $k > 1$
Weibull survival	$S(t) = e^{-(\lambda t)^k}$	
Weibull mean	$\mu = \Gamma(1 + 1/k)/\lambda$	
Weibull MLE ($\hat{\lambda}$)	$\hat{\lambda}^k = n / \sum t_i^k$	Closed form
Log-Normal hazard	$h(t) = \phi(z)/[\sigma t\, \Phi(-z)]$, $z = (\ln t - \mu)/\sigma$	Unimodal
Log-Normal mean	$\mu_T = e^{\mu + \sigma^2/2}$	
Kaplan–Meier	$\hat{S}(t) = \prod_{t_j \le t}(1 - d_j/n_j)$	Product-limit
Nelson–Aalen	$\hat{H}(t) = \sum_{t_j \le t} d_j/n_j$	
Greenwood variance	$\hat{V}[\hat{S}] = \hat{S}^2 \sum_{t_j \le t} d_j/[n_j(n_j - d_j)]$	
Cox PH hazard	$h(t\|\mathbf{x}) = h_0(t) \exp(\boldsymbol{\beta}^\top \mathbf{x})$	Semi-param.

(continued)

Quantity	Formula	Notes
Cox partial likelihood	$\mathcal{L}(\boldsymbol{\beta}) = \prod_j \exp(\boldsymbol{\beta}^\top \mathbf{x}_{(j)}) / \sum_{i \in \mathcal{R}(t_j)} \exp(\boldsymbol{\beta}^\top \mathbf{x}_i)$	
Breslow estimator	$\hat{H}_0(t) = \sum_{t_j \le t} d_j / \sum_{i \in \mathcal{R}(t_j)} e^{\hat{\boldsymbol{\beta}}^\top \mathbf{x}_i}$	
AFT model	$\ln T = \boldsymbol{\beta}^\top \mathbf{x} + \sigma \epsilon$	Accelerated time
Frailty survival	$S(t \mid \mathbf{x}) = \mathcal{L}_Z(H_0(t) e^{\boldsymbol{\beta}^\top \mathbf{x}})$	Gamma: $(1 + \theta H_0 e^{\boldsymbol{\beta}^\top \mathbf{x}})^{-1/\theta}$
Log-rank statistic	$\chi^2 = (\sum_j (O_{1j} - E_{1j}))^2 / \sum_j \hat{V}_j$	2-group

Numerical Tour: Part 6 — Fatigue Life Prediction Pipeline

A complete fatigue life prediction for a welded steel connection in a railway bridge, from load spectrum to probabilistic lifetime distribution.

Setting. Bridge main girder, welded connection detail. Material: S355 structural steel. Fatigue category (Eurocode 3): FAT 71 (endurance limit $\Delta\sigma_D = 52.3$ MPa at $N_D = 5 \times 10^6$ cycles; slope $m = 3$ in S-N curve). Annual traffic load spectrum from WIM (Weigh-in-Motion) data: 10^7 cycles per year across 5 stress range bins:

Stress range $\Delta\sigma_i$ (MPa)	Cycles/year n_i	$\Delta\sigma_i > \Delta\sigma_D$?
20	5.0×10^6	No (below endurance)
35	3.0×10^6	No
55	1.2×10^6	Yes
75	5.0×10^5	Yes
100	8.0×10^4	Yes

Step 1: Miner damage per year. Cycles below $\Delta\sigma_D$ are non-damaging (omitted). For the three active bins, the S-N curve gives $N_i = C\Delta\sigma_i^{-3}$ with $C = \Delta\sigma_D^3 N_D = 52.3^3 \times 5 \times 10^6 = 7.16 \times 10^{10}$ (MPa^3·cycles):

$\Delta\sigma_i$	n_i	$N_i = C/\Delta\sigma_i^3$	$d_i = n_i/N_i$
55	1.2×10^6	$7.16 \times 10^{10}/55^3 = 4.31 \times 10^5$	2.785
Wait – $N_i = 7.16 \times 10^{10}/166375 = 4.30 \times 10^5$			
$\to d_i = 1.2 \times 10^6/4.30 \times 10^5$			2.791

Let me redo cleanly. $C = 52.3^3 \times 5 \times 10^6$. $52.3^3 = 143\,000$ (approx). $C = 1.43 \times 10^5 \times 5 \times 10^6 = 7.15 \times 10^{11}$ MPa^3·cycles.

$\Delta\sigma_i$	$\Delta\sigma_i^3$	$N_i = C/\Delta\sigma_i^3$	n_i/year	d_i
55	1.664×10^5	$7.15 \times 10^{11}/1.664 \times 10^5 = 4.30 \times 10^6$	1.2×10^6	0.279
75	4.219×10^5	$7.15 \times 10^{11}/4.219 \times 10^5 = 1.70 \times 10^6$	5.0×10^5	0.295
100	1.000×10^6	$7.15 \times 10^{11}/1.000 \times 10^6 = 7.15 \times 10^5$	8.0×10^4	0.112
			Annual Miner damage D_{yr}:	**0.686**

Step 2: Deterministic design life. Miner failure at $D = 1$: $N_{\text{life}} = 1/D_{\text{yr}} = 1/0.686 = 1.46$ years. With Miner critical damage $D_c = 0.5$ (Eurocode partial factor for variable amplitude): $N_{\text{life}}^{\text{safe}} = 0.5/0.686 = 0.73$ years. This is far too short — the connection requires redesign or fatigue upgrading.

Step 3: Redesign via weld improvement. Post-weld treatment (hammer peening) improves the fatigue category from FAT 71 to FAT 90, shifting $\Delta\sigma_D$ to 66.4 MPa and C to $7.15 \times 10^{11} \times (90/71)^3 = 1.46 \times 10^{12}$.

Recalculated D_{yr} (only $\Delta\sigma > 66.4$ MPa contribute): bins at 75 and 100 MPa only.

$\Delta\sigma_i$	N_i (new C)	n_i	d_i
75	$1.46 \times 10^{12}/4.219 \times 10^5 = 3.46 \times 10^6$	5.0×10^5	0.145
100	$1.46 \times 10^{12}/10^6 = 1.46 \times 10^6$	8.0×10^4	0.055
		New D_{yr}:	0.200

New design life (at $D_c = 0.5$): $0.5/0.200 = 2.5$ years. Still insufficient for a 60-year bridge life.

Step 4: Probabilistic fatigue assessment. Using the IIW probabilistic approach: $\ln N \sim$

$\mathcal{N}(\ln C - m\ln\Delta\sigma, \sigma^2_{\ln N})$ with $\sigma_{\ln N} = 0.25$ (typical for welded details). The annual damage $D_{\text{yr}} = \sum_i n_i/N_i$ is log-normally distributed with:

$$\mu_{\ln D} = \ln\left(\sum_i n_i/N_i^{\text{median}}\right) - \sigma^2_{\ln N}/2 \approx \ln(0.200) - 0.031 = -1.658,$$
$$\sigma_{\ln D} \approx 0.25 \text{ (dominated by S-N scatter)}.$$

Target: $\Pr(D_{60\,\text{yr}} \geq 1.0) \leq 10^{-4}$ (code requirement for bridges). $D_{60} \sim \text{LN}(60\mu_{\ln D}, 60^{0.5} \times 0.25)$? No: $D_{60} = 60D_{\text{yr}}$, so $\ln D_{60} \sim \mathcal{N}(60\mu_{\ln D} + \ln 60, \sigma_{\ln N})$. Probability of Miner failure: $\Pr(\ln D_{60} \geq 0) = \Phi\left(\frac{-60\times(-1.658)-\ln 60}{0.25}\right) = \Phi\left(\frac{99.48-4.094}{0.25}\right) = \Phi(381) \approx 0$. Good: the 60-year failure probability is effectively zero under FAT 90 and the given spectrum. The bridge connection passes.

Formula Reference: Part 6 — Fatigue and Degradation

Quantity	Formula	Notes
Basquin's law	$N = CS^{-m}$	S-N curve
Coffin–Manson	$\Delta\varepsilon_p/2 = \varepsilon'_f(2N)^c$	Low-cycle
Goodman (mean stress)	$S_a/S_e + \sigma_m/\sigma_{UTS} = 1$	Conservative
Stress concentration	$K_f = 1 + (K_t - 1)/(1 + a/r)$	Notch effect
Miner damage	$D = \sum n_i/N_i$	Failure at $D \geq 1$
LEFM SIF range	$\Delta K = Y\Delta\sigma\sqrt{\pi a}$	
Paris' law	$da/dN = C_P(\Delta K)^m$	Region II
Cycles to failure	$N_f = \int_{a_0}^{a_c} da/[C_P(Y\Delta\sigma\sqrt{\pi a})^m]$	
For $m \neq 2$	$N_f = 2[(a_0^{1-m/2} - a_c^{1-m/2})]/[(m-2)C_P(Y\Delta\sigma)^m\pi^{m/2}]$	
Critical crack	$a_c = (K_{Ic}/(Y\sigma_{\max}))^2/\pi$	LEFM limit
Walker eq.	$da/dN = C_P[\Delta K/(1-R)^{1-\gamma}]^m$	Stress ratio
Gamma process mean	$\mathbb{E}[X(t)] = \alpha(t)/\beta$	
Gamma process CDF	$\Pr(T \leq t) = 1 - F_\Gamma(\ell; \alpha(t), \beta)$	FPT
Wiener FPT	$T \sim \text{IG}(\ell/\mu, \ell^2/\sigma^2)$	IG distribution
IG density	$f_T(t) = \frac{\ell}{\sqrt{2\pi\sigma^2 t^3}}\exp(-(\ell-\mu t)^2/(2\sigma^2 t))$	
Wiener FPT CDF	$\Pr(T \leq t) = \Phi((\mu t-\ell)/(\sigma\sqrt{t})) + e^{2\mu\ell/\sigma^2}\Phi((-\mu t-\ell)/(\sigma\sqrt{t}))$	
Remaining useful life	$\text{RUL} \sim \text{IG}((\ell - x_m)/\hat{\mu}, (\ell - x_m)^2/\sigma^2)$	Online update

Numerical Tour: Part 7 — Optimising a Fleet Maintenance Programme

Complete worked example optimising a multi-component maintenance programme for a fleet of 50 diesel-electric locomotives.

System. Each locomotive has 3 critical components:

Component	Distribution	μ (days)	C_f (repair)	C_p (planned)
Traction motor (TM)	Weibull, $k = 2.5$	180	\$4500	\$600
Air compressor (AC)	Weibull, $k = 1.8$	120	\$1200	\$250
Wheel bearings (WB)	Weibull, $k = 3.2$	240	\$900	\$200

Fleet size: $N = 50$ locomotives. Operating days: 300/year (maintenance windows available on non-operating days).

Step 1: Individual optimal replacement ages. For each component, solve the age replacement optimality condition:

$$h(T_p^*) \int_0^{T_p^*} S(t)\,dt - F(T_p^*) = \frac{C_p}{C_f - C_p}.$$

Traction motor ($k = 2.5$, $\mu = 180$, $\lambda = \Gamma(1.4)/180 = 0.887/180 = 0.00493$): RHS $= 600/(4500 - 600) = 0.154$. Numerically: $T_p^*(\text{TM}) \approx 142$ days. Long-run cost rate $g(142) = (600 + 3900 \times F(142))/\int_0^{142} S(t)dt$. $F(142) = 1 - e^{-(0.00493\times 142)^{2.5}} = 1 - e^{-0.700^{2.5}} = 1 - e^{-0.408} = 0.335$. $\int_0^{142} S(t)dt \approx 142 \times \bar{S} \approx 142 \times (1 - 0.335/2) = 142 \times 0.833 = 118$ days. $g_{\text{TM}} = (600 + 3900 \times 0.335)/118 = (600 + 1307)/118 = \$16.2$/day. Compare to failure-only: $g(\infty) = C_f/\mu = 4500/180 = \$25.0$/day. Savings: 35%.

Air compressor ($k = 1.8$, $\mu = 120$): $T_p^*(\text{AC}) \approx 98$ days, $g_{\text{AC}} = \$8.9$/day.

Wheel bearings ($k = 3.2$, $\mu = 240$): $T_p^*(\text{WB}) \approx 195$ days, $g_{\text{WB}} = \$3.8$/day.

Step 2: Fleet-level cost aggregation. Total maintenance cost rate per locomotive: $g_{\text{total}} = g_{\text{TM}} + g_{\text{AC}} + g_{\text{WB}} = 16.2 + 8.9 + 3.8 = \$28.9$/day. Fleet annual cost: $50 \times 300 \times 28.9 = \$433,500$/year.

Step 3: Opportunistic maintenance. When TM is replaced (every ≈ 142 days on average), inspect AC and WB. If their ages exceed opportunistic thresholds, replace them simultaneously (setup cost saved: \$300/component per joint maintenance event).

Opportunistic threshold for WB (current age a_{WB} at TM replacement): Marginal cost of replacing WB at a_{WB} vs. waiting: replace opportunistically if

$$g_{\text{WB}}(a_{\text{WB}}) > \frac{C_p^{\text{WB}} - \Delta c_{\text{setup}}}{c_{\text{additional}}} \quad \Rightarrow \quad a^*_{\text{WB,opp}} \approx 155 \text{ days}.$$

If WB is older than 155 days when TM is replaced, replace both together. Simulation over 2000 maintenance cycles: 28% of TM replacements trigger WB opportunistic replacement. Annual saving from opportunistic maintenance: $\approx \$18,000$ (4.2% of fleet maintenance budget).

Step 4: Competing risks analysis. For each locomotive, time to first component failure: $T_{\text{min}} = \min(T_{\text{TM}}, T_{\text{AC}}, T_{\text{WB}})$. Under independence:

$$S_{\text{min}}(t) = S_{\text{TM}}(T_p^*) \cdot S_{\text{AC}}(T_p^*) \cdot S_{\text{WB}}(T_p^*).$$

With planned replacement, the effective per-component survival (between planned replacements) has much higher reliability than the baseline Weibull — the effective fleet MTBF

(any component failure) under the optimised PM programme: ≈ 38 days (vs. 12 days for failure-only).

Step 5: Spare parts planning. Expected annual replacements per locomotive:

- TM: $300/142 \times [F(142) + 1] \approx 300/142 \times 1.335 = 2.82$ replacements/year (including planned).
- AC: $300/98 \times 1.310 = 4.01$ replacements/year.
- WB: $300/195 \times 1.175 = 1.81$ replacements/year.

Fleet annual demand: TM: $50 \times 2.82 = 141$; AC: $50 \times 4.01 = 201$; WB: $50 \times 1.81 = 91$ units. Using a (s, Q) inventory with safety stock $z_{0.95}\sqrt{12 \times \text{monthly demand} \times \text{lead time var}}$ for each spare.

Formula Reference: Part 7 — Renewal and Maintenance

Quantity	Formula	Notes
Renewal equation	$M(t) = F(t) + \int_0^t M(t-s)\,dF(s)$	Fundamental
ERT	$M(t)/t \to 1/\mu$ as $t \to \infty$	Asymptotic
Asymptotic expansion	$M(t) \approx t/\mu + (\sigma^2 - \mu^2)/(2\mu^2)$	2nd order
Variance of $N(t)$	$\text{Var}[N(t)] \approx \sigma^2 t/\mu^3$	Large t
Asymptotic availability	$A(\infty) = \mu_U/(\mu_U + \mu_V)$	Alternating
Age replacement rate	$g(T_p) = (C_p + (C_f - C_p)F(T_p))/\int_0^{T_p} S(t)\,dt$	
Optimal age condition	$h(T_p^*)\int_0^{T_p^*} S(t)\,dt - F(T_p^*) = C_p/(C_f - C_p)$	
Block replacement rate	$g_B(T_p) = (C_p + c_f M(T_p))/T_p$	
Renewal reward theorem	$\lim_{t\to\infty}$ total reward$/t = \mathbb{E}[R]/\mathbb{E}[X]$	
Kijima I virtual age	$v_k = q \cdot t_k$	
Kijima II virtual age	$v_k = v_{k-1} + q(t_k - t_{k-1})$	
Virtual age hazard	$h_{k+1}(t) = h_0(v_k + t)$	
Geometric process	$\mathbb{E}[X_k] = a^{k-1}\mu$	$a > 1$: deterioration
Crow-AMSAA intensity	$\rho(t) = \lambda\beta t^{\beta-1}$	Growth: $\beta < 1$
C-A MLE shape	$\hat{\beta} = n/\sum_{i=1}^n \ln(T/t_i)$	
Cause-specific hazard	$h_k(t) = \Pr(t < T \le t + dt, C = k \mid T > t)/dt$	
CIF	$F_k(t) = \int_0^t S(s^-)h_k(s)\,ds$	

Quantity	Formula	Notes
Aalen-Johansen	$\hat{F}_k(t) = \sum_{t_j \le t} \hat{S}(t_j^-) d_{kj}/n_j$	
Subdistrib. hazard	$\tilde{h}_k(t) = -d\ln(1-F_k(t))/dt$	Fine-Gray

Numerical Tour: Part 8 — System Reliability for a Redundant Power Supply

Full quantitative analysis of an uninterruptible power supply (UPS) system for a data centre.

Architecture. The UPS system provides 2N redundancy: two independent power paths (A and B), each with: utility grid input (G), static bypass switch (S), rectifier (R), battery bank (B), and inverter (I). Normal operation: both paths deliver power in parallel ($N+1$ mode). On failure of any component in Path A: Path B carries full load (degraded mode). System fails only if both paths fail simultaneously or if the static bypass fails (single path).

Simplified series-parallel model.

$$\phi(\mathbf{x}) = \phi_{\text{SB}}(\mathbf{x}) \cdot [\phi_A(\mathbf{x}) \cup \phi_B(\mathbf{x})],$$

where $\phi_A = G_A R_A B_A I_A$ (series within path A), $\phi_B = G_B R_B B_B I_B$, and ϕ_{SB} is the static bypass (always in series).

Component failure rates (IEC 62395 and field data):

Component	λ (failures/year)	$p = e^{-\lambda \times 1}$
Grid connection (G)	0.50	0.607
Rectifier (R)	0.10	0.905
Battery bank (B)	0.08	0.923
Inverter (I)	0.12	0.887
Static bypass (S)	0.02	0.980

Path A reliability: $p_A = 0.607 \times 0.905 \times 0.923 \times 0.887 = 0.607 \times 0.905 = 0.549$; $\times 0.923 = 0.507$; $\times 0.887 = 0.450$.

System reliability (without bypass): $R_{\text{paths}} = 1 - (1 - p_A)^2 = 1 - (0.550)^2 = 1 - 0.302 = 0.698$. $R_s = p_S \times R_{\text{paths}} = 0.980 \times 0.698 = 0.684$.

The 1-year system availability of 68.4% is unacceptably low. Analysis reveals the grid connection (G, $\lambda = 0.50$) dominates.

Redesign with dual utility feeds. Replace single G with two independent utility feeds (G_1, G_2) in parallel within each path, each with $\lambda = 0.50$: $p_{G_{\text{par}}} = 1 - (1 - 0.607)^2 = 1 - 0.154 = 0.846$.

New path reliability: $p_A^{\text{new}} = 0.846 \times 0.905 \times 0.923 \times 0.887 = 0.846 \times 0.905 = 0.766$; $\times 0.923 = 0.707$; $\times 0.887 = 0.627$.

$R_{\text{paths}}^{\text{new}} = 1 - (1 - 0.627)^2 = 1 - 0.139 = 0.861$. $R_s^{\text{new}} = 0.980 \times 0.861 = 0.844$. Better, but still below the 99.999% ("5 nines") availability target.

2N+1 architecture. Add a third path C (partial redundancy): system works if at least 1 of 3 paths works AND the bypass is functional. $R_{\text{paths}}^{2N+1} = 1 - (1 - 0.627)^3 = 1 - 0.051 = 0.949$. $R_s^{2N+1} = 0.980 \times 0.949 = 0.930$. Still insufficient.

Root cause: batteries. With MTBF of $1/0.08 = 12.5$ years ($p_B = 0.923$), batteries are not the bottleneck. The grid feeds are. With dual-feed redundancy AND $2N$ path architecture AND annual battery maintenance (reducing effective λ_B to 0.02):

$p_B^{\text{maint}} = e^{-0.02} = 0.980$. $p_A^{\text{final}} = 0.846 \times 0.905 \times 0.980 \times 0.887 = 0.666$. $R_{\text{paths}}^{\text{final}} = 1 - (1 - 0.666)^2 = 1 - 0.112 = 0.888$. $R_s^{\text{final}} = 0.980 \times 0.888 = 0.870$.

Importance measures (final design): $I_{\text{SB}}^{FV} = (1 - p_S)/(1 - R_s) = 0.020/0.130 = 0.154$. $I_{G_{\text{par}},\text{pathA}}^{B} = R_s(G = 1) - R_s(G = 0) = \ldots \approx 0.035$.

The static bypass remains the single highest-importance component (15.4% of system un-

reliability as a singleton). The recommendation: upgrade the static bypass to $\lambda_S = 0.005$ (solid-state with self-test): $p_S = 0.995$. New $R_s = 0.995 \times 0.888 = 0.884$.

For 5-nines availability ($R_s \geq 0.99999$), the analysis shows this single-site 2N architecture is insufficient regardless of component reliabilities. The path to 5-nines is geographic redundancy (two independent data centres) — a system-of-systems architecture outside the component-level model of this volume.

CCF analysis. With $\beta = 0.05$ for the rectifiers (common design, same vendor): CCF failure rate of both rectifiers: $\lambda_{\text{CCF},R} = 0.05 \times 0.10 = 0.005$/year. CCF contribution to path failure: $P_{\text{CCF}} = 0.005 \times 1\,\text{year} = 0.005$. New path reliability: $p_A^{\text{CCF}} = p_A^{\text{final}} \times (1 - P_{\text{CCF}}/p_A) \approx 0.666 - 0.005 = 0.661$. Minor impact in this case because the grid connection dominates.

Formula Reference: Part 8 — System Reliability

Quantity	Formula	Notes
Series reliability	$R_s = \prod_{i=1}^{n} p_i$	
Parallel reliability	$R_s = 1 - \prod_i (1 - p_i)$	
k-out-of-n	$R_{k/n} = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}$	Identical p
Series MTTF (exp.)	$\text{MTTF} = 1/\sum_i \lambda_i$	
Parallel MTTF (exp.)	$\text{MTTF} = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k\lambda} \binom{n}{k}^{-1} \sum_{\lvert S\rvert=k} (\sum_{i\in S} \lambda_i)^{-1}$	
Birnbaum importance	$I_i^B = R_s(1_i, \mathbf{p}) - R_s(0_i, \mathbf{p}) = \partial R_s / \partial p_i$	
Criticality importance	$I_i^C = I_i^B (1 - p_i)/(1 - R_s)$	
Fussell–Vesely	$I_i^{FV} = \sum_{K_j \ni i} P(K_j)/(1 - R_s)$	
RAW	$\text{RAW}_i = (1 - R_s(0_i))/(1 - R_s)$	
RRW	$\text{RRW}_i = (1 - R_s)/(1 - R_s(1_i))$	
β-factor PFD	$P_{\text{CCF}} \approx \beta \lambda_{DU} \tau / 2$	
1oo2 PFD	$\text{PFD}_{1oo2} \approx (\lambda_{DU}\tau/2)^2 + \beta\lambda_{DU}\tau/2$	With CCF
SIL 1	$\text{PFD} \in [10^{-2}, 10^{-1})$	IEC 61508
SIL 2	$\text{PFD} \in [10^{-3}, 10^{-2})$	
SIL 3	$\text{PFD} \in [10^{-4}, 10^{-3})$	
RAP (series-parallel)	$\max \prod_j (1 - (1 - p_j)^{n_j})$ s.t. $\sum c_j n_j \leq C$	
RAP optimal age cond.	Solved by DP or Lagrangian relaxation	
Cold standby MTTF	$S_{\text{cs}}(t) = e^{-\lambda t}(1 + \lambda t)$	Primary + 1 cold spare

Quantity	Formula	Notes
Warm standby	$S_{\text{ws}}(t) = e^{-\lambda t} + \lambda/(\lambda - \lambda_{sb})(e^{-\lambda_{sb} t} - e^{-\lambda t})$	

Deep Dive: Decision Analysis and Value of Information in Reliability

Expected value of perfect information (EVPI). Before committing to a maintenance action a under uncertain component reliability θ, the decision-maker faces:

$$\text{EV}(\text{best action without info}) = \max_a \mathbb{E}_\theta[U(a, \theta)].$$

With perfect information about θ:

$$\text{EV}(\text{perfect info}) = \mathbb{E}_\theta[\max_a U(a, \theta)].$$

The EVPI (willingness to pay for perfect information):

$$\text{EVPI} = \text{EV}(\text{perfect info}) - \text{EV}(\text{best action}).$$

If EVPI is small, additional testing or monitoring is not worth the investment; if large, further information gathering is valuable.

Expected value of sample information (EVSI). With imperfect information $\mathbf{y}$ (e.g. n test results), the EVSI is:

$$\text{EVSI}(n) = \mathbb{E}_{\mathbf{y}}\left[\max_a \mathbb{E}_{\theta|\mathbf{y}}[U(a, \theta)]\right] - \max_a \mathbb{E}_\theta[U(a, \theta)].$$

EVSI is non-decreasing in n and bounded above by EVPI. The optimal sample size n^* solves $\text{EVSI}(n^* + 1) - \text{EVSI}(n^*) \leq c_{\text{sample}}$ (marginal benefit equals marginal cost).

Value of health monitoring (VoHM). In condition monitoring, the VoHM is the difference in expected total lifecycle cost between the CBM policy (using degradation data) and the best age/block replacement policy (without data):

$$\text{VoHM} = g_{\text{best PM}}(T_p^*) - g_{\text{CBM}}(d^*),$$

per unit time. This is the maximum willingness to pay for sensor installation and monitoring infrastructure. For Gamma degradation processes with moderate variance ($\text{CV}_X \approx 0.5$), VoHM typically represents 15–30% of the PM-only maintenance cost.

Example — EVPI for turbine blade inspection programme

Prior on blade Weibull shape: $k \sim \text{Uniform}[1.5, 3.5]$ (uncertainty about wear-out rate). Two actions: a_1 = inspect every 500 FH (expensive but safe), a_2 = inspect every 1000 FH (cheap but riskier for IFR blades).
Expected costs: $\mathbb{E}[U(a_1, k)] = \$180$/FH (constant, not k-dependent). $U(a_2, k) = \$90 + \$500 \times (F_k(1000) - F_k(500))$/FH — higher for larger k (more failures between inspections).
$\mathbb{E}_k[U(a_2)] = \$90 + \$500 \times \mathbb{E}_k[F_k(1000) - F_k(500)] = \$90 + \$500 \times 0.18 = \$180$/FH.
Under prior uncertainty: both actions have the same expected cost! But with per-

fect info: if $k < 2.1$ (DFR/mild IFR), a_2 is cheaper; if $k > 2.1$, a_1 is cheaper. EVPI = $\$500 \times \int_{2.1}^{3.5} [F_k(1000) - F_k(500) - 0.18] \tfrac{1}{2}\, dk \approx \$8$/FH. This EVPI of \$8/FH justifies spending up to \$8/FH (approximately \$70,000/year for a fleet of 1000 flight hours/year) on an improved inspection programme that reduces k-uncertainty.

Probabilistic Bounds and Approximation Theory in Reliability

Bounds on Failure Probability

First-order bounds (Bonferroni/union bounds). For a system with K failure modes (minimal cut sets $\{K_1, \dots, K_m\}$), the system failure probability satisfies:

$$P_f \le \sum_{j=1}^{m} P(K_j) \quad \text{(union bound, always valid)},$$
$$P_f \ge \max_j P(K_j) \quad \text{(trivial lower bound)},$$
$$P_f \ge \sum_j P(K_j) - \sum_{j<k} P(K_j \cap K_k) \quad \text{(Bonferroni lower bound)}.$$

The union bound is tight when cut set probabilities are small (rare event regime $P(K_j) \ll 1$).

Chernoff bounds for fleet reliability. For a fleet of n independent identical components with failure probability p each, the probability that the number of failures exceeds $\mu + \epsilon$ (where $\mu = np$) is bounded by Chernoff:

$$\Pr(D \ge \mu + \epsilon) \le \exp\left(-\frac{\epsilon^2}{2(\mu + \epsilon/3)}\right).$$

This exponential bound is essential for fleet operators who must guarantee that total failures D stay within budget with high probability.

Chebyshev bound for system availability. If $A(t)$ is the system availability at time t with $\mathbb{E}[A] = \bar{A}$ and $\mathrm{Var}[A] = \sigma_A^2$:

$$\Pr(A(t) \le \bar{A} - \epsilon) \le \frac{\sigma_A^2}{\epsilon^2}.$$

For an availability target of $\bar{A} - \epsilon = 0.99$: if $\bar{A} = 0.995$ and $\sigma_A = 0.01$, $\Pr(A < 0.99) \le (0.01)^2/(0.005)^2 = 4$ — the Chebyshev bound is too loose here; tighter bounds (e.g. using the beta distribution for A) give $\Pr(A < 0.99) \approx 0.02$.

FORM and SORM for Reliability Index

First-order reliability method (FORM). For a failure criterion $g(\mathbf{X}) \le 0$ where $\mathbf{X} = (X_1, \dots, X_n)$ are random variables, the failure probability is:

$$P_f = \Pr(g(\mathbf{X}) \le 0) = \int_{g(\mathbf{x}) \le 0} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}.$$

After transforming to standard normal space ($\mathbf{U} = T(\mathbf{X})$), the limit state $G(\mathbf{U}) = g(T^{-1}(\mathbf{U})) = 0$ becomes a surface in $\mathbb{R}^n$. The most probable point of failure (MPP or design point) $\mathbf{u}^*$ is the point on the limit state surface closest to the origin: $\mathbf{u}^* = \arg\min_{\mathbf{u}:\, G(\mathbf{u})=0} \|\mathbf{u}\|$. The FORM approximation:

$$P_f^{\text{FORM}} = \Phi(-\beta_{\text{HL}}), \quad \beta_{\text{HL}} = \|\mathbf{u}^*\|$$

(Hasofer–Lind reliability index). For a single failure mode: $\beta_{\text{HL}} = (\mu_g)/\sigma_g$ (ratio of mean safety margin to its standard deviation) for linear g and normal variables.

Second-order reliability method (SORM). The FORM approximation linearises $G = 0$ at $\mathbf{u}^*$. SORM fits a paraboloid: correcting for the curvature κ_i (principal curvatures of the limit state surface at $\mathbf{u}^*$):

$$P_f^{\text{SORM}} \approx \Phi(-\beta) \prod_{i=1}^{n-1} (1 + \beta\kappa_i)^{-1/2} \quad \text{(Breitung formula)}.$$

SORM improves on FORM accuracy by $O(\beta^{-2})$ relative to the exact P_f. SORM is essential when the limit state is highly nonlinear or when P_f is very small ($< 10^{-4}$).

Example — FORM for combined axial and bending load on a shaft

Shaft strength $R \sim \mathcal{N}(\mu_R = 400\,\text{MPa}, \sigma_R = 40\,\text{MPa})$. Applied stress $S = S_1 + S_2$ where $S_1 \sim \mathcal{N}(150, 20^2)$ (axial) and $S_2 \sim \mathcal{N}(80, 15^2)$ (bending), independent. Limit state: $g = R - S_1 - S_2$.
$\mu_g = 400 - 150 - 80 = 170$ MPa. $\sigma_g = \sqrt{40^2 + 20^2 + 15^2} = \sqrt{1600 + 400 + 225} = \sqrt{2225} = 47.2$ MPa.
$\beta_{\text{HL}} = 170/47.2 = 3.60$. $P_f^{\text{FORM}} = \Phi(-3.60) = 1.59 \times 10^{-4}$.
Since g is linear, FORM is exact here. The sensitivity factors: $\alpha_R = 40/47.2 = 0.847$ (strength is the dominant uncertainty), $\alpha_{S_1} = 20/47.2 = 0.424$, $\alpha_{S_2} = 15/47.2 = 0.318$. The gradient vector ∇G shows that reducing σ_R (improving strength consistency) is the most effective way to increase β.

Reliability index and safety factors. The safety factor $\text{SF} = \mu_R/\mu_S$ is related to β_{HL} by:

$$\text{SF} = \frac{1 + \beta_{\text{HL}}\text{CV}_R}{1 - \beta_{\text{HL}}\text{CV}_S},$$

where $\text{CV} = \sigma/\mu$. For $\text{CV}_R = 0.10$, $\text{CV}_S = 0.15$, and target $\beta = 3.5$ (structural engineering target $P_f \approx 2 \times 10^{-4}$): $\text{SF} = (1 + 0.35)/(1 - 0.525) = 1.35/0.475 = 2.84$. This provides a probabilistic calibration of deterministic safety factors.

Exercises and Problems — Part 5

The following problems range from computational (marked **[C]**) to theoretical (**[T]**) to applied (**[A]**). Starred problems **[⋆]** require more extended work.

5.1. **[C]** A component has hazard function $h(t) = 0.002 + 0.0001t$ (linear increasing). (a) Find the cumulative hazard $H(t)$ and survival function $S(t)$. (b) Compute the probability of surviving to $t = 50$ h. (c) Find the median failure time (solve $S(t_{0.5}) = 0.5$ numerically). (d) Show that this distribution is IFR and find the mean residual life $e(0) = \mathbb{E}[T]$.

5.2. **[C]** Given the following 10 failure times (hours): 12, 25, 38, 47, 63, 82, 105, 140, 198, 275. Fit a Weibull distribution by: (a) probability plotting (log-log transformation); (b) MLE (solve the score equations numerically). Compare the two estimates of k and λ.

5.3. **[T]** Prove that the exponential distribution is the unique continuous distribution satisfying the memoryless property $\Pr(T > s + t | T > s) = \Pr(T > t)$. *Hint*: show that $S(s+t) = S(s)S(t)$ implies $S(t) = e^{-\lambda t}$.

5.4. **[C]** Six components are tested with the following (time, status) observations: $(50, 1), (80, 0), (110, 1), (130$ (a) Compute the Kaplan–Meier estimate at all failure times. (b) Compute the Nelson–Aalen estimate and the corresponding $\tilde{S}(t) = e^{-\hat{H}(t)}$. (c) Compare KM and NA estimates: which is larger? Why?

5.5. **[A]** A reliability engineer fits a log-normal distribution to bearing failure times and obtains $\hat{\mu} = 7.2$ and $\hat{\sigma} = 0.6$ (in log-hours). (a) Find the 10th-percentile life (B10 life) and the median. (b) Compute the hazard at $t = 500$ h and $t = 2000$ h. Show that the log-normal hazard is unimodal and find the mode of $h(t)$. (c) The component is replaced at $T_p = 1500$ h. What fraction of replacements will be preventive (unit still working)?

5.6. **[T]** For the Weibull distribution with shape $k > 1$ (IFR), prove that the optimal preventive replacement age $T_p^* < \infty$ exists and is unique when $C_p < C_f$. *Hint*: show that $g'(T_p) < 0$ for small T_p and $g'(T_p) > 0$ for large T_p by analysing the sign of $h(T_p)\int_0^{T_p} S(t)\,dt - F(T_p)$.

5.7. **[C]** A Cox PH model is fitted with two covariates: x_1 = age (standardised), x_2 = maintenance quality index. The estimated coefficients are $\hat{\beta}_1 = 0.35$, $\hat{\beta}_2 = -0.48$. (a) Interpret the hazard ratios. (b) For a unit with $x_1 = 1.5$, $x_2 = -0.5$, compute the relative risk compared to the baseline ($\mathbf{x} = \mathbf{0}$). (c) If the baseline survival at 1000 h is $\hat{S}_0(1000) = 0.72$, what is the predicted survival for this unit?

5.8. **[⋆][T]** Show that the Information matrix $\mathcal{I}(\boldsymbol{\beta}) = -\mathbb{E}[\partial^2 \ell_{\text{partial}}/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top]$ for the Cox partial likelihood equals $\sum_j \text{Var}_{\boldsymbol{\beta}}[\mathbf{X} | \text{failure at } t_j]$, the sum of covariate variances in the risk set at each failure time. Conclude that $\mathcal{I}(\boldsymbol{\beta})$ is always positive definite (ensuring the partial likelihood is strictly concave).

5.9. **[A]** An analyst observes 20 component failures and fits both an exponential and a Weibull distribution. The log-likelihoods are $\ell_E = -87.4$ and $\ell_W = -84.1$. (a) Perform a likelihood ratio test for $H_0 : k = 1$ vs $H_1 : k \neq 1$. What is the test statistic and p-value? (b) Compute AIC and BIC for both models. Which model is preferred? (c) The Weibull MLE gives $\hat{k} = 1.42$. What does this imply about the optimal maintenance strategy?

5.10. [⋆][C] *Competing risks for transformer failures.* A fleet of 80 transformers can fail due to: insulation breakdown (cause 1, $\hat{\lambda}_1 = 0.04$/yr), overheating (cause 2, $\hat{\lambda}_2 = 0.02$/yr), or mechanical failure (cause 3, $\hat{\lambda}_3 = 0.01$/yr). (a) Compute the cause-specific hazards and the overall survival function (assuming independence). (b) Compute the CIF for each cause at $t = 10$ years. Verify that $F_1(10) + F_2(10) + F_3(10) + S(10) = 1$. (c) If a reliability improvement programme eliminates cause 2 failures ($\hat{\lambda}_2 \to 0$), what is the new 20-year survival probability? (d) Explain why this calculation is only valid under independence of causes.

Exercises and Problems — Part 6

6.1. [C] A steel alloy has Basquin parameters $C = 8 \times 10^{12}$ MPa^3·cycles and $m = 3$. (a) Find the endurance limit if the S-N curve becomes horizontal at $N_D = 2 \times 10^6$ cycles. (b) A component experiences a constant amplitude loading of $\Delta\sigma = 250$ MPa. How many cycles until fatigue failure? (c) A second component experiences $\Delta\sigma = 350$ MPa for 5×10^4 cycles, then $\Delta\sigma = 200$ MPa. Using Miner's rule, how many additional cycles at 200 MPa are allowed?

6.2. [T] Derive the number of cycles to failure N_f by integrating Paris' law $da/dN = C_P(\Delta K)^m$ for a general m, with $\Delta K = Y\Delta\sigma\sqrt{\pi a}$ and constant $Y, \Delta\sigma$. Show that for $m = 2$ the result involves $\ln(a_c/a_0)$, and for $m \neq 2$ it involves power laws in a_0 and a_c.

6.3. [C] A pressure vessel made of aluminium alloy has: initial crack $a_0 = 1$ mm, critical crack $a_c = 20$ mm (at $K_{Ic} = 30$ MPa$\sqrt{\text{m}}$), geometry factor $Y = 1.0$, Paris constants $C_P = 2 \times 10^{-11}$, $m = 3.2$, stress range $\Delta\sigma = 80$ MPa. (a) Verify a_c using $a_c = (K_{Ic}/(Y\sigma_{\max}))^2/\pi$ with $\sigma_{\max} = \Delta\sigma$ (R=0). (b) Compute N_f analytically. (c) What initial crack a_0^* would give $N_f = 10^6$ cycles?

6.4. [A] A gearbox tooth has a Weibull failure distribution ($k = 2.8$, $\mu = 15000$ cycles). Degradation is modelled as a Gamma process with mean $\mu_X(t) = 0.05t$ (wear depth in μm) and rate parameter $\beta = 10$. The failure threshold is $\ell = 300\,\mu$m. (a) Compute the mean and variance of the Gamma process at $t = 5000$ cycles. (b) Compute the probability $\Pr(X(5000) > 300)$ using the Gamma CDF. (c) Compare with the Weibull survival $S(5000)$. Which model predicts higher reliability?

6.5. [T] For the Wiener process $X(t) = \mu t + \sigma W_t$, show that the first passage time $T = \inf\{t : X(t) \geq \ell\}$ has the Inverse Gaussian distribution with mean $\mathbb{E}[T] = \ell/\mu$ and variance $\text{Var}[T] = \ell\sigma^2/\mu^3$. *Hint*: use the Optional Stopping Theorem applied to the martingales $X(t) - \mu t$ and $(X(t) - \mu t)^2 - \sigma^2 t$.

6.6. [C] A structural steel beam is subjected to a random load spectrum modelled as a narrow-band Gaussian process with RMS stress $\sigma_\sigma = 30$ MPa and mean zero-crossing rate $\nu_0 = 2$ Hz. Basquin parameters: $m = 4$, $C = 5 \times 10^{18}$ MPa^4·cycles. (a) Using the narrow-band approximation, compute the expected damage per second: $\mathbb{E}[\dot{D}] = \nu_0 C^{-1}(2\sigma_\sigma^2)^{m/2}\Gamma(1 + m/2)$. (b) Estimate the expected time to failure (when $D = 1$). (c) If the stress level increases by 10% ($\sigma_\sigma \to 33$ MPa), what is the percentage reduction in expected life? (Note: life scales as σ_σ^{-m}.)

6.7. [⋆][A] *Probabilistic fracture mechanics.* For a turbine disc with uncertain initial crack size $a_0 \sim \text{LN}(\mu_{\ln a} = \ln(0.5\,\text{mm}), \sigma_{\ln a} = 0.5)$ and deterministic Paris law ($C_P = 5 \times 10^{-12}$, $m = 3.5$, $Y = 1.12$, $\Delta\sigma = 120$ MPa, $K_{Ic} = 55$ MPa$\sqrt{\text{m}}$): (a) Show that $a_c = (K_{Ic}/(Y\Delta\sigma))^2/\pi$. (b) For fixed a_0, derive the closed-form $N_f(a_0)$. (c) Compute $\Pr(N_f \leq N^*)$ for $N^* = 10^5$ cycles by integrating over the a_0 distribution (numerically).

(d) Find the a_0 value that gives $N_f = N^*$ (the critical initial crack size).

6.8. **[T]** The Kachanov–Rabotnov damage evolution $\dot{\omega} = A\sigma^r/(1-\omega)^k$ has the closed-form solution $\omega(t) = 1 - [1-(k+1)A\sigma^r t]^{1/(k+1)}$. (a) Verify this by substitution. (b) Find the failure time t_f (when $\omega = 1$). (c) Show that $t_f = 1/[(k+1)A\sigma^r]$. (d) For a sequence of two constant-stress periods, derive the Miner-analogous cumulative damage rule implied by the Kachanov model. Is it linear or nonlinear in n_i/N_i?

Exercises and Problems — Part 7

7.1. **[C]** The inter-renewal times of a process follow an Erlang-2 distribution ($\lambda = 0.1$, $k = 2$, so $\mu = 2/\lambda = 20$). (a) Compute $M(t)$ at $t = 10, 20, 50, 100$ using the exact formula for Erlang-2 renewals. (b) Compare with the asymptotic approximation $M(t) \approx t/\mu + (\sigma^2 - \mu^2)/(2\mu^2)$ where $\sigma^2 = 2/\lambda^2 = 200$. (c) Plot (sketch) the renewal function and identify where the oscillations are most pronounced.

7.2. **[A]** A pump has Weibull failure distribution ($k = 2.2$, $\mu = 800$ h). Planned replacement cost $C_p = \$400$; failure replacement cost $C_f = \$2500$. (a) Verify that the hazard is IFR and that preventive replacement is beneficial. (b) Find the optimal replacement age T_p^* numerically (solve the first-order condition). (c) Compute the long-run cost rate $g(T_p^*)$ and compare to the failure-only cost $g(\infty) = C_f/\mu$. (d) How sensitive is T_p^* to a 20% increase in C_p?

7.3. **[T]** Prove the Renewal Reward Theorem: if $\{(X_i, R_i)\}$ are i.i.d. with $\mathbb{E}[|R_i|] < \infty$ and $\mathbb{E}[X_i] = \mu < \infty$, then $\frac{1}{t}\sum_{k=1}^{N(t)} R_k \to \frac{\mathbb{E}[R]}{\mu}$ almost surely as $t \to \infty$. *Hint*: Write the sum as $\frac{N(t)}{t} \cdot \frac{1}{N(t)}\sum_{k=1}^{N(t)} R_k$ and apply the SLLN to each factor using the ERT.

7.4. **[C]** For the Kijima Model I with Weibull baseline ($k = 2.5$, $\lambda = 0.001$) and repair effectiveness $q = 0.6$: (a) Simulate (or calculate analytically) the virtual ages after 5 repair cycles if the actual failure times are $t_1 = 100$, $t_2 = 180$, $t_3 = 250$, $t_4 = 310$, $t_5 = 390$ h. (b) Compute the conditional hazard just after each repair. (c) Compare with the AGAN ($q = 0$) and ABAO ($q = 1$) models. How much faster does ABAO deteriorate compared to AGAN by the 5th cycle?

7.5. **[A]** Analyse the following 12 failure times (hours) from a repairable pump using the Crow–AMSAA model: 45, 82, 131, 198, 253, 310, 378, 451, 530, 615, 710, 820. Observation ends at $T = 900$ h. (a) Apply the Laplace trend test. Is there evidence of reliability growth or deterioration? (b) Compute $\hat{\beta}$ and $\hat{\lambda}$ by MLE. (c) Predict the number of failures in the next 100 h (from $t = 900$ to 1000 h). (d) At what cumulative time will the 15th failure be expected to occur?

7.6. **[T]** In the competing risks model with K independent exponential failure modes ($T_k \sim \text{Exp}(\lambda_k)$), show that: (a) $T_{\min} = \min_k T_k \sim \text{Exp}(\sum_k \lambda_k)$. (b) The cause of failure is $C = \arg\min_k T_k$ with $\Pr(C = k) = \lambda_k / \sum_j \lambda_j$, independently of $T_{\min}$. (c) The CIF for cause k is $F_k(t) = (\lambda_k/\sum_j \lambda_j)(1 - e^{-\sum_j \lambda_j t})$. (d) Why does eliminating cause k (setting $\lambda_k \to 0$) change the distribution of $T_{\min}$ but *not* the marginal distribution of T_j for $j \neq k$?

7.7. **[⋆][A]** *Optimal maintenance for a deteriorating fleet.* A fleet of 30 aircraft engines follows a geometric process with $a = 1.15$ (15% reduction in mean time between overhauls each cycle), initial MTBO $\mu_1 = 3000$ h, and repair cost $c_k = \$50{,}000 \cdot 1.08^{k-1}$ (8% cost escalation per cycle). Engine replacement cost: $C_R = \$2{,}000{,}000$. (a) Compute the expected total time between overhauls for $N = 1, 2, 3, 5, 8$ repairs before replacement.

(b) Compute the long-run cost rate $g(N)$ for each. (c) Find the optimal N^*. (d) How does N^* change if the cost escalation rate drops from 8% to 4%?

7.8. **[T]** Derive Greenwood's formula for $\mathrm{Var}[\hat{S}(t)]$ in the Kaplan–Meier estimator. *Hint*: Write $\ln \hat{S}(t) = \sum_{t_j \le t} \ln(1 - d_j/n_j)$. Apply the delta method: $\mathrm{Var}[\ln \hat{S}] \approx \sum_j \mathrm{Var}[d_j/n_j]/(1 - d_j/n_j)^2$, and use $\mathrm{Var}[d_j] \approx n_j (d_j/n_j)(1 - d_j/n_j)$ (binomial approximation) to obtain $\mathrm{Var}[\ln \hat{S}] \approx \sum_j d_j / [n_j (n_j - d_j)]$. Apply the delta method again to get $\mathrm{Var}[\hat{S}]$.

Exercises and Problems — Part 8

8.1. **[C]** A 3-component system has the structure function $\phi(x_1, x_2, x_3) = x_1 (x_2 + x_3 - x_2 x_3)$ (component 1 in series with a parallel pair). With $p_1 = 0.9$, $p_2 = 0.8$, $p_3 = 0.75$: (a) Compute the system reliability. (b) Find all minimal path sets and minimal cut sets. (c) Compute Birnbaum, FV, and criticality importance for each component. (d) Which component should be upgraded first to maximise ΔR_s per unit cost if upgrade costs are $c_1 = 3$, $c_2 = 2$, $c_3 = 1$ and each upgrade raises the component reliability by 0.05?

8.2. **[T]** Prove that the Birnbaum importance $I_i^B = \partial R_s / \partial p_i$ satisfies the bounds $0 \le I_i^B \le 1$ for any coherent structure function. *Hint*: use the path set representation $R_s = 1 - \prod_{K \in \mathcal{K}} (1 - \prod_{i \in K} (1 - p_i))$ and differentiate; show monotonicity in p_i.

8.3. **[A]** For a 4-component series system with $p_i = [0.95, 0.90, 0.85, 0.92]$ and failure rates $\lambda_i = [-\ln p_i / T]$ (1-year): (a) Compute the system reliability and MTTF (exponential model). (b) Compute all four Birnbaum importances. (c) If a budget allows one component to be made perfectly reliable ($p_i \to 1$), which component gives the maximum ΔR_s? (d) If instead the budget allows reducing each λ_i by 20%, which component's improvement gives the most ΔR_s?

8.4. **[C]** A fault tree has the top event $T = (A + B)(C + D)$ (two parallel groups in series). Basic event probabilities: $p_A = 0.03$, $p_B = 0.05$, $p_C = 0.02$, $p_D = 0.04$. (a) Find all minimal cut sets. (b) Compute $P(\text{top event})$ exactly. (c) Apply the rare event approximation and compare. (d) Add a common cause failure for A and B with β-factor $\beta = 0.1$: recompute $P(\text{top event})$.

8.5. **[T]** In the Markov model for a 2-component parallel system (λ, μ): states are $\{2, 1, 0\}$ (number of working components). (a) Write the transition rate matrix Q. (b) Solve for the steady-state distribution $\pi Q = 0$, $\sum \pi = 1$. (c) Show that the asymptotic unavailability is $U = \pi_0 = \lambda^2/(\lambda^2 + 2\lambda\mu + \mu^2/\text{something})$... derive the exact formula. (d) For $\lambda = 0.01$/h and $\mu = 0.5$/h, compute U and compare with the static approximation $U \approx (\lambda/\mu)^2 \cdot$ (correction factor).

8.6. **[⋆][A]** *Redundancy allocation for a 4-stage missile guidance system.* Series system, 4 subsystems, $p_j = [0.80, 0.85, 0.90, 0.75]$, $c_j = [3, 2, 4, 1]$ (cost units), total budget $B = 14$. (a) Set up the dynamic programming recursion. (b) Build the complete DP table for $W = 0$ to 14 and stages $j = 4$ down to $j = 1$. (c) Find the optimal allocation $(n_1^*, n_2^*, n_3^*, n_4^*)$ and the maximum system reliability. (d) Verify by checking all feasible solutions with cost ≤ 14 (there are only $\binom{14}{4}$ at most).

8.7. **[A]** An IEC 61508 SIL 2 safety function requires $\mathrm{PFD}_{\text{avg}} \in [10^{-3}, 10^{-2})$. The safety function uses a 1oo2 voting architecture with $\lambda_{\text{DU}} = 5 \times 10^{-6}$/h per channel, proof-test interval τ, and $\beta = 0.08$. (a) Compute PFD as a function of τ. (b) Find the maximum τ that keeps PFD $< 10^{-2}$ (SIL 2 upper limit). (c) Find the minimum τ that keeps PFD $> 10^{-3}$ (SIL 2 lower limit — if PFD is too low, SIL 3 applies and additional requirements kick

in). (d) What is the cost-optimal τ if proof tests cost \$5000 each and SIL 2 compliance is required?

8.8. **[⋆][T]** *Monotone systems and the IFRA closure theorem.* (a) Define a coherent binary monotone system. (b) State and prove that if $T_1, \dots, T_n$ are i.i.d. with distribution $F \in$ IFR, then the system lifetime $T_s = \phi(T_1, \dots, T_n)$ is also IFR. *Hint*: show that for a series system, the sum of IFR hazards is IFR; for a parallel system, use the fact that $S_s(t) = 1 - (1 - S(t))^n$ is log-concave. Then use the modular decomposition theorem to extend to general coherent systems.

Bibliographic Notes and Further Reading

Part 5 — Survival Analysis

The foundational text is **Kalbfleisch & Prentice** (2002), *The Statistical Analysis of Failure Time Data* (2nd ed., Wiley), covering censoring, partial likelihood, and nonparametric estimation rigorously. **Fleming & Harrington** (1991), *Counting Processes and Survival Analysis* (Wiley), provides the martingale theory underlying the Nelson–Aalen and Cox estimators. **Lawless** (2003), *Statistical Models and Methods for Lifetime Data* (2nd ed., Wiley) covers AFT and log-normal models comprehensively. Frailty models: **Hougaard** (2000), *Analysis of Multivariate Survival Data* (Springer). The random survival forest: **Ishwaran & Kogalur** (2008), Ann. Appl. Statist. 2(3), 841–860. Bayesian survival with Beta processes: **Hjort** (1990), Ann. Statist. 18, 1259–1294. Joint degradation-survival models: **Tsiatis & Davidian** (2004), Biometrika, 91(3), 447–458.

Part 6 — Fatigue, Fracture, and Degradation

Fatigue of materials: **Suresh** (1998), *Fatigue of Materials* (2nd ed., Cambridge UP). Fracture mechanics: **Anderson** (2017), *Fracture Mechanics: Fundamentals and Applications* (4th ed., CRC). The J-integral and EPFM: **Rice** (1968), J. Appl. Mech. 35, 379–386. Paris' law: **Paris & Erdogan** (1963), J. Basic Engng. 85, 528–534. Miner's rule: **Miner** (1945), J. Appl. Mech. 12, A159–A164. Dirlik spectral fatigue: **Dirlik** (1985), PhD thesis, University of Warwick. Gamma process degradation: **van Noortwijk** (2009), Reliability Engng. & System Safety, 94, 2–21. Wiener process and IG distribution: **Whitmore & Schenkelberg** (1997), Lifetime Data Analysis, 3, 27–45. Continuum damage mechanics: **Lemaitre & Chaboche** (1990), *Mechanics of Solid Materials* (Cambridge UP). Probabilistic fracture mechanics: **Provan** (ed., 1987), *Probabilistic Fracture Mechanics and Reliability* (Martinus Nijhoff).

Part 7 — Renewal Theory and Maintenance

Classical reliability theory: **Barlow & Proschan** (1996), *Mathematical Theory of Reliability* (SIAM Classics). Renewal theory: **Feller** (1971), *Introduction to Probability Theory and Its Applications*, Vol. II (Wiley), Chapters XI–XIII. Maintenance optimisation: **Jardine & Tsang** (2013), *Maintenance, Replacement and Reliability* (2nd ed., CRC). Kijima virtual age models: **Kijima** (1989), J. Appl. Prob. 26, 89–102. Geometric process: **Lam** (1988), Adv. Appl. Prob. 20, 479–482. Crow–AMSAA reliability growth: **Crow** (1974), in *Proc. IEEE Annual Reliability & Maintainability Symposium*. Competing risks identifiability: **Tsiatis** (1975), Proc. Natl. Acad. Sci. 72, 20–22. Fine–Gray model: **Fine & Gray** (1999), JASA 94, 496–509. Aalen–Johansen estimator: **Aalen & Johansen** (1978), Scand. J. Statist. 5, 141–150.

Part 8 — System Reliability

Structure functions and importance measures: **Barlow & Proschan** (1996) (cited above). Fault tree handbook: **Vesely et al.** (1981), *Fault Tree Handbook* (NUREG-0492, US NRC). Common-cause failure: **Fleming & Mosleh** (1985), EPRI NP-3967; **Mosleh et al.** (1998), NUREG/CR-5485. Binary decision diagrams for reliability: **Rauzy** (1993), Reliability Engng. & System Safety, 40, 45–56. Dynamic fault trees: **Dugan et al.** (1992), IEEE Trans. Reliability, 41, 363–377. Redundancy allocation: **Ramirez-Marquez & Coit** (2004), Reliability Engng. & System Safety, 83, 341–349. Multi-state systems and UGF: **Lisnianski & Levitin** (2003), *Multi-State System Reliability* (World Scientific). Reliability-centred maintenance: **Moubray** (1997), *Reliability-Centred Maintenance* (2nd ed., Butterworth-Heinemann).

IEC 61508 functional safety: International Electrotechnical Commission, *Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems*, Parts 1–7 (2010).

Software and Computational Tools

- **R**: packages `survival` (KM, Cox, AFT), `frailtypack` (frailty models), `cmprsk` (competing risks, Fine–Gray), `ReliabilityTheory` (structure functions).
- **Python**: `lifelines` (KM, Cox, AFT, Nelson–Aalen), `reliability` (parametric MLE, degradation, competing risks), `scikit-survival` (machine learning for survival data).
- **Julia**: `Survival.jl` (high-performance survival), `Turing.jl` (Bayesian inference for all models in this volume).
- **MATLAB**: Statistics and Machine Learning Toolbox (parametric MLE, Weibull), Reliability Toolbox (RBD, FTA).
- **OpenTURNS** (open source C++/Python): FORM/SORM, importance sampling, copula-based dependence for CCF modelling.
- **Industry PRA tools**: CAFTA (Electrical Power Research Institute), RiskSpectrum (Lloyd's Register), SAPHIRE (Idaho National Laboratory) for nuclear PRA.

Computational Notes: Algorithms and Implementation

Algorithm 1: Weibull MLE via Newton–Raphson

Given n uncensored failure times $\{t_i\}$ (extension to censored data: add $-\lambda^k c_j^k$ terms from censored observations c_j):

Score equations:

$$U_k = \frac{\partial \ell}{\partial k} = \frac{n}{k} + \sum_{i=1}^{n} \ln(\lambda t_i)[1 - (\lambda t_i)^k],$$
$$U_\lambda = \frac{\partial \ell}{\partial \lambda} = \frac{nk}{\lambda} - k\lambda^{k-1} \sum_{i=1}^{n} t_i^k.$$

From $U_\lambda = 0$: $\hat{\lambda}(k) = (n/\sum t_i^k)^{1/k}$. Substituting into $U_k = 0$ gives a single equation in k, solved by Newton iteration: $k_{m+1} = k_m - U_k(k_m)/U_k'(k_m)$. Converges in 5–10 iterations from $k_0 = 1$.

Standard errors: From the observed Fisher information matrix $\hat{\mathbf{I}} = -\nabla^2 \ell\big|_{\hat{\theta}}$:

$$\mathrm{SE}(\hat{k}) = \sqrt{[\hat{\mathbf{I}}^{-1}]_{11}}, \quad \mathrm{SE}(\hat{\lambda}) = \sqrt{[\hat{\mathbf{I}}^{-1}]_{22}}.$$

95% Wald CIs: $\hat{k} \pm 1.96\,\mathrm{SE}(\hat{k})$ (use log transformation for k and λ to ensure positivity: CI for $\ln k$ then exponentiate).

Algorithm 2: Kaplan–Meier Estimator with Log-Log Confidence Bands

Input: $(t_i, \delta_i)_{i=1}^n$ pairs. **Output**: step function $\hat{S}(t)$ with 95% log-log confidence bands.

Steps:

1. Sort by t_i; at each distinct failure time t_j: compute n_j (at risk) and d_j (failures).
2. $\hat{S}(t_j) = \hat{S}(t_{j-1})(1 - d_j/n_j)$.
3. Greenwood variance accumulator: $V_j = V_{j-1} + d_j/[n_j(n_j - d_j)]$.
4. Log-log CI at t_j: $\hat{S}(t_j)^{\exp(\pm 1.96\sqrt{V_j}/|\ln \hat{S}(t_j)|)}$.

Ties: if d_j observations fail at the same time as c_j censored observations, treat censored observations as occurring after failures (standard convention).

Median survival: smallest t^* such that $\hat{S}(t^*) \le 0.5$. CI for median via inversion of the pointwise CI for $\hat{S}(t)$.

Algorithm 3: Cox Partial Likelihood Optimisation

Partial log-likelihood:

$$\ell(\boldsymbol{\beta}) = \sum_{j:\,\delta_j=1} \left[\boldsymbol{\beta}^\top \mathbf{x}_j - \ln \sum_{k:\,t_k \ge t_j} e^{\boldsymbol{\beta}^\top \mathbf{x}_k} \right].$$

Score: $U(\boldsymbol{\beta}) = \sum_{j:\,\delta_j=1} [\mathbf{x}_j - \bar{\mathbf{x}}(\boldsymbol{\beta}, t_j)]$ where $\bar{\mathbf{x}} = \sum_{k \in \mathcal{R}(t_j)} w_k \mathbf{x}_k$ and $w_k = e^{\boldsymbol{\beta}^\top \mathbf{x}_k} / \sum_{l \in \mathcal{R}} e^{\boldsymbol{\beta}^\top \mathbf{x}_l}$.

Hessian: $H(\boldsymbol{\beta}) = -\sum_j \mathrm{Var}_w[\mathbf{X} \mid \mathcal{R}(t_j)]$ (weighted covariance matrix in risk set; negative definite $\Rightarrow$ partial likelihood is concave).

Newton–Raphson: $\boldsymbol{\beta}_{m+1} = \boldsymbol{\beta}_m - H^{-1}(\boldsymbol{\beta}_m)U(\boldsymbol{\beta}_m)$. Initialise at $\boldsymbol{\beta}_0 = \mathbf{0}$. Convergence in 5–20 iterations. For p large: use L-BFGS with analytic gradient.

Breslow baseline: After convergence: $\hat{H}_0(t) = \sum_{j:\, t_j \le t} d_j / \sum_{k \in \mathcal{R}(t_j)} e^{\hat{\boldsymbol{\beta}}^\top \mathbf{x}_k}$.

Algorithm 4: Gamma Process MLE via L-BFGS-B

Parameters: (c, η, β) with $\Delta\alpha_{ij} = c(t_j^\eta - t_{j-1}^\eta)$.

Log-likelihood:

$$\ell = \sum_{i=1}^{n}\sum_{j=1}^{m}\left[(\Delta\alpha_{ij} - 1)\ln \Delta x_{ij} - \Delta x_{ij}\beta + \Delta\alpha_{ij}\ln\beta - \ln\Gamma(\Delta\alpha_{ij})\right].$$

Gradient components (require the digamma function $\psi = \Gamma'/\Gamma$):

$$\begin{aligned}
\frac{\partial \ell}{\partial c} &= \sum_{ij}(\ln \Delta x_{ij} - \ln\beta - \psi(\Delta\alpha_{ij}))(t_j^\eta - t_{j-1}^\eta),\\
\frac{\partial \ell}{\partial \eta} &= \sum_{ij}(\ln \Delta x_{ij} - \ln\beta - \psi(\Delta\alpha_{ij}))\cdot c(t_j^\eta \ln t_j - t_{j-1}^\eta \ln t_{j-1}),\\
\frac{\partial \ell}{\partial \beta} &= \sum_{ij}\left(\frac{\Delta\alpha_{ij}}{\beta} - \Delta x_{ij}\right).
\end{aligned}$$

Implementation tip: Parameterise as $(\ln c, \ln\eta, \ln\beta)$ to enforce positivity; L-BFGS-B converges in 30–100 function evaluations from a warm start at $(c_0, \eta_0, \beta_0) = (\bar{\Delta x}/\bar{\Delta t}, 1, 1/\bar{\Delta x})$.

Algorithm 5: Renewal Function Computation via FFT

For discrete inter-renewal distribution $\{f_k\}_{k=1}^K$ (probabilities on $\{1, \dots, K\}$), the renewal mass function $m_n = M(n)$ satisfies:

$$m_n = \sum_{k=1}^{n} f_k(1 + m_{n-k}), \quad m_0 = 0.$$

Direct recursion: $O(n^2)$. FFT acceleration:

1. Embed in array of length $2K$: $\tilde{f} = [0, f_1, \dots, f_K, 0, \dots, 0]$.
2. Compute $\hat{f} = \mathrm{FFT}(\tilde{f})$.
3. $\hat{m} = \hat{f}/(1 - \hat{f})$ (elementwise, in frequency domain).
4. $m = \mathrm{IFFT}(\hat{m})$.

Complexity: $O(K \log K)$. Accuracy: exact for discrete distributions; for continuous F, discretise with step δ to obtain $O(\delta^2)$ approximation error.

Algorithm 6: Dynamic Programming for the Redundancy Allocation Problem

Input: k stages, component reliabilities $\{p_j\}$, costs $\{c_j\}$, budget B.

DP recursion: $f_j(w)$ = maximum reliability for stages $j, \dots, k$ with budget w.

$$f_j(w) = \max_{1 \le n \le \lfloor w/c_j \rfloor} [1 - (1-p_j)^n] \cdot f_{j+1}(w - n\,c_j), \quad f_{k+1}(w) = 1.$$

Complexity: $O(k \cdot B/c_{\min} \cdot B/c_{\max})$ — polynomial in B but pseudo-polynomial in the input encoding. For $k = 4$, $B = 20$, this is $\le 4 \times 20 \times 20 = 1600$ operations (trivially fast).

Traceback: Store $n_j^*(w) = \arg\max_n[\dots]$ at each (j, w); recover optimal allocation by forward pass: $w_1 = B$, $n_j^* = n_j^*(w_j)$, $w_{j+1} = w_j - n_j^* c_j$.

Algorithm 7: Monte Carlo Simulation of Age Replacement Policy

Input: Weibull parameters $(\hat{k}, \hat{\lambda})$, replacement age T_p, costs C_p, C_f, horizon $N_{\text{sim}} = 10^5$ cycles.

Vectorised implementation (Python/NumPy sketch):

Generate N_{sim} Weibull samples: `T_fail = scipy.stats.weibull_min.rvs(k, scale=1/lam, size=N)`.

Replacement time per cycle: `T_cycle = np.minimum(T_fail, T_p)`.

Cost per cycle: `cost = np.where(T_fail <= T_p, C_f, C_p)`.

Long-run cost rate: `g_hat = cost.mean() / T_cycle.mean()`.

Standard error: delta method, $\widehat{\text{SE}}[g] = g\sqrt{(\sigma_C/\mu_C)^2 + (\sigma_T/\mu_T)^2 - 2\rho_{CT}\sigma_C\sigma_T/(\mu_C\mu_T)}$.

Grid search for T_p^*: evaluate $\hat{g}(T_p)$ on a grid $T_p \in [0.1\mu, 3\mu]$ with 100 points; fit a cubic spline and find the minimum. Alternatively, use `scipy.optimize.minimize_scalar` with `method='golden'` for the theoretical $g(T_p)$.

Algorithm 8: Importance Sampling for Rare Failure Events

For failure probability $P_f = \Pr(D \ge \ell) \ll 1$, standard MC variance $= P_f(1-P_f)/N$ requires $N \gg 1/P_f$.

Exponential twisting (Cramér tilt):

1. Find $\theta^* = \arg\min_\theta[\ln M_D(\theta) - \theta\ell]$ where $M_D(\theta) = \mathbb{E}[e^{\theta D}]$.
2. Sample from tilted distribution $\tilde{F}(x) = \int_0^x e^{\theta^* u - \ln M_D(\theta^*)}\, dF(u)$.
3. Likelihood ratio weight: $L = e^{-\theta^* D^{(k)} + \ln M_D(\theta^*)}$.
4. Estimate: $\hat{P}_f = \frac{1}{N}\sum_{k=1}^N \mathbf{1}[D^{(k)} \ge \ell] \cdot L^{(k)}$.

Variance reduction: $\text{Var}[\hat{P}_f^{\text{IS}}] = O(P_f^2)$ vs. $\text{Var}[\hat{P}_f^{\text{MC}}] = O(P_f/N)$ — exponential improvement for small P_f. Effective sample size $\approx P_f \cdot N$ vs. $P_f^2 \cdot N$ required for crude MC.

Glossary of Key Terms and Symbols

Term / Symbol	Definition and context
Survival Analysis (Part 5)	
T	Non-negative random variable representing time to failure.

(continued)

Term / Symbol	Definition and context
$F(t)$	Cumulative distribution function: $\Pr(T \le t)$.
$S(t)$	Survival function: $1 - F(t) = \Pr(T > t)$.
$f(t)$	Probability density: $f(t) = F'(t)$.
$h(t)$	Hazard (failure rate) function: $f(t)/S(t)$. Instantaneous conditional failure rate. Units: time^{-1}.
$H(t)$	Cumulative hazard: $H(t) = -\ln S(t) = \int_0^t h(u)\,du$.
$e(t)$	Mean residual life at age t: $\mathbb{E}[T - t \mid T > t] = \int_t^\infty S(u)\,du / S(t)$.
IFR	Increasing Failure Rate: $h(t)$ non-decreasing. Wear-out behaviour.
DFR	Decreasing Failure Rate: $h(t)$ non-increasing. Infant mortality.
IFRA	Increasing Failure Rate Average: $H(t)/t$ non-decreasing. Closed under coherent system formation.
BT	Bathtub hazard: decreasing then increasing. Composite of infant mortality, useful life, and wear-out.
δ_i	Censoring indicator: $\delta_i = 1$ if failure observed, $\delta_i = 0$ if right-censored.
n_j, d_j	Risk set size and number of failures at ordered event time t_j (KM/Nelson–Aalen).
KM	Kaplan–Meier product-limit estimator: $\hat{S}(t) = \prod_{t_j \le t}(1 - d_j/n_j)$. Nonparametric MLE of $S(t)$ under censoring.
NA	Nelson–Aalen estimator: $\hat{H}(t) = \sum_{t_j \le t} d_j/n_j$. Nonparametric MLE of $H(t)$.
Cox PH	Cox Proportional Hazards model: $h(t\mid\mathbf{x}) = h_0(t)\exp(\boldsymbol{\beta}^\top\mathbf{x})$. Semi-parametric regression for survival data.
Partial likelihood	Likelihood for $\boldsymbol{\beta}$ in Cox model, profiling out the baseline hazard $h_0(t)$.
AFT	Accelerated Failure Time model: $\ln T = \boldsymbol{\beta}^\top\mathbf{x} + \sigma\epsilon$. Covariates stretch/compress the time axis.
Frailty	Unobserved random effect $Z > 0$ multiplying the hazard: $h(t\mid Z) = Z h_0(t) e^{\boldsymbol{\beta}^\top\mathbf{x}}$. Models unobserved heterogeneity.
Schoenfeld residuals	Difference between covariate at failure time and its risk-set expectation. Diagnostic for PH assumption.
Fatigue, Fracture, and Degradation (Part 6)	
S-N curve	Stress–Number of cycles curve (Wöhler curve). Maps constant-amplitude stress S to cycles to failure N.
Basquin's law	$N = CS^{-m}$. Power-law S-N relationship. Parameters: C (intercept), m (slope in log-log).
S_e	Endurance limit (fatigue limit). Stress below which fatigue failure does not occur in ferrous metals.
m	Basquin/Paris exponent. S-N slope ($m = 3$–8); Paris crack growth exponent ($m = 2$–4).
Miner's rule	$D = \sum n_i/N_i$. Linear cumulative damage. Failure when $D \ge D_c$ (often $D_c = 1$).
Rainflow counting	Algorithm (Matsuishi–Endo) to extract closed stress cycles from irregular load histories.
K, ΔK	Stress intensity factor and its range. $K = Y\sigma\sqrt{\pi a}$. Governs crack tip stress field in LEFM.

(continued)

Term / Symbol	Definition and context
K_{Ic}	Plane-strain fracture toughness. Critical K at which fast fracture occurs. Material property.
Paris' law	$da/dN = C_P(\Delta K)^m$. Fatigue crack growth rate in Region II (LEFM regime).
POD	Probability of Detection for nondestructive testing. $P_{\text{det}}(a)$: function of crack size a.
$a_{90/95}$	Crack size detectable with 90% probability at 95% confidence. NDT capability metric.
Gamma process	Stochastic process with independent, non-negative, stationary increments. Natural model for monotone degradation.
IG distribution	Inverse Gaussian distribution. First-passage time distribution for Wiener processes with positive drift.
RUL	Remaining Useful Life. Time from current state to failure threshold crossing. Key condition monitoring output.
FPT	First Passage Time: $T = \inf\{t : X(t) \geq \ell\}$.
Fokker–Planck	PDE governing the time evolution of the probability density of a diffusion process. Used to compute FPT densities.
Renewal Theory and Maintenance (Part 7)	
Renewal process	$\{N(t)\}$: count of i.i.d. inter-renewal events in $[0, t]$. Models repeated failures and repairs.
$M(t)$	Renewal function: $\mathbb{E}[N(t)]$. Satisfies the renewal equation $M = F + F * M$.
ERT	Elementary Renewal Theorem: $M(t)/t \to 1/\mu$ as $t \to \infty$.
RRT	Renewal Reward Theorem: long-run reward rate = $\mathbb{E}[R]/\mathbb{E}[X]$.
Age replacement	Replace at failure or at age T_p (whichever first). Optimal for IFR distributions.
Block replacement	Replace all at fixed times kT_p; minimal repair at failures between.
T_p^*	Optimal preventive replacement age. Minimises long-run cost rate $g(T_p)$.
CBM	Condition-Based Maintenance. Triggers replacement based on observed degradation level, not calendar time.
Virtual age	Kijima's imperfect repair model: after repair, effective age is $v_k < t_k$ (partial rejuvenation).
q	Repair effectiveness (Kijima). $q = 0$: perfect repair (AGAN). $q = 1$: minimal repair (ABAO).
Geometric process	$\{X_k\}$ with $\mathbb{E}[X_k] = a^{k-1}\mu$. Models deteriorating repairable systems with $a > 1$.
NHPP	Non-Homogeneous Poisson Process. Models failures of repairable systems under minimal repair with changing rate $\rho(t)$.
Crow–AMSAA	NHPP power law model: $\Lambda(t) = \lambda t^\beta$. Used for reliability growth analysis.
CIF	Cumulative Incidence Function. $F_k(t) = \Pr(T \leq t, C = k)$. Direct estimate of cause-specific failure probability under competing risks.
Fine–Gray	Regression model for the subdistribution hazard in competing risks. Directly models the CIF.

(continued)

Term / Symbol	Definition and context
Aalen–Johansen	Nonparametric estimator of the CIF matrix for competing risks.
System Reliability (Part 8)	
Structure function	$\phi : \{0,1\}^n \to \{0,1\}$. Maps component states to system state.
Minimal path set	Minimal set of working components that guarantees system function.
Minimal cut set	Minimal set of failed components that causes system failure.
I_i^B	Birnbaum importance: $\partial R_s/\partial p_i = R_s(1_i) - R_s(0_i)$.
I_i^{FV}	Fussell–Vesely importance: fraction of system unreliability from MCS containing i.
RAW	Risk Achievement Worth: factor increase in system unreliability if component i always fails.
RRW	Risk Reduction Worth: factor decrease in system unreliability if component i never fails.
CCF	Common-Cause Failure: dependent failure of multiple components from a shared root cause.
β-factor	Fraction of total failure rate attributable to CCF events affecting all components in a group.
PFD	Probability of Failure on Demand: average probability that a safety system fails when demanded.
SIL	Safety Integrity Level (IEC 61508). Discrete classification of safety system reliability requirements.
RAP	Redundancy Allocation Problem. Optimal assignment of redundant components to system stages subject to cost/weight constraints.
BDD	Binary Decision Diagram. Compact DAG representation of Boolean structure functions enabling efficient reliability computation.
RCM	Reliability-Centred Maintenance. Systematic process for determining optimal maintenance tasks based on failure mode consequences.
FMEA	Failure Mode and Effects Analysis. Bottom-up systematic identification of failure modes, causes, and effects.
FMECA	FMEA with Criticality Analysis. Adds quantitative ranking of failure mode risk via RPN or criticality number.
PRA	Probabilistic Risk Assessment. Quantitative system-level risk analysis integrating ETA and FTA.
DFT	Dynamic Fault Tree. Extends static FTA with dynamic gates (spare, priority AND, sequence enforcing).

Open Problems and Research Frontiers

The following open problems at the frontier of reliability theory are intended to motivate advanced readers and point toward active research directions.

1. Identifiability in imperfect repair models. The virtual age of Kijima's model is unobservable; the repair effectiveness q must be inferred from inter-failure times alone. Under what conditions is q identifiable from a single repair history? For Weibull baseline distributions, the Fisher information for q approaches zero as $k \to 1$ (exponential) — the MLE is consistent but its convergence rate slows. *Open*: sharp minimax lower bounds on the estimation error for q as a function of the number of repair cycles n and the Weibull shape k.

2. Non-proportional hazards in high dimensions. The Cox model assumes hazard proportionality — an assumption routinely violated in practice (e.g. treatment effects that diminish over time). While time-varying coefficients and stratification address specific violations, a general nonparametric test for the proportionality assumption in the high-dimensional ($p \gg n$) regime remains underdeveloped. *Open*: derive the asymptotic distribution of a goodness-of-fit test for the Cox model when the number of covariates p grows with n.

3. Optimal stopping for FPT models with model uncertainty. Given a Wiener degradation process $X(t)$ with uncertain drift μ (estimated online), the optimal replacement rule is a time-varying threshold $x^*(t)$ derived from the HJB equation. When μ is uncertain and estimated from noisy observations (Bayesian learning), the problem becomes a *learning while acting* (exploration-exploitation) stochastic control problem. *Open*: characterise the value of information and the optimal balance between continued monitoring and preemptive replacement under parameter uncertainty.

4. Network reliability computation for large-scale systems. All-terminal network reliability ($\Pr(\text{all pairs connected})$) is #P-hard to compute exactly. For sparse random graphs (Erdős–Rényi, power-law degree distributions), mean-field approximations give accuracy $O(1/n)$. For heterogeneous networks (varying edge reliabilities, multiple terminals), Monte Carlo methods with importance sampling achieve variance reduction of $O(e^{-n})$ near the percolation threshold. *Open*: provably efficient importance sampling for general network reliability problems near phase transitions.

5. Joint modelling of longitudinal degradation and failure in high dimensions. The joint model of Section 5.5 (degradation trajectory + survival) is well-developed for scalar degradation. When K degradation biomarkers are monitored simultaneously ($K = 100$–1000 in digital twin applications), the model becomes a high-dimensional longitudinal regression coupled with a survival outcome. Existing approaches use dimensionality reduction (PCA on the degradation trajectories) followed by standard joint modelling, but this discards interaction information. *Open*: scalable Bayesian joint models for high-dimensional degradation processes with good RUL prediction accuracy.

6. Distributional robustness of maintenance policies. Section 7.2 derived optimal maintenance policies under known parametric distributions. When the failure time distribution is only known to lie in a Wasserstein ball around an estimated distribution, what is the optimal *distributionally robust* maintenance policy? For age replacement, the worst-case distribution in a Wasserstein ball $\mathcal{W}_p(\hat{F}, \epsilon)$ may have a heavier tail than $\hat{F}$, making the optimal age policy more conservative. *Open*: characterise the optimal robust age and CBM policies and their dependence on the ambiguity radius ϵ.

7. Physics-informed neural networks for degradation modelling. Recent work (Raissi et al., 2019) uses neural networks constrained by physical PDEs (Fokker–Planck, Paris' law) as loss function terms. For degradation modelling, a physics-informed NN (PINN) could

simultaneously fit observed degradation data and satisfy the underlying SDE. This offers the flexibility of deep learning without sacrificing physical interpretability. *Open*: theoretical guarantees (convergence, generalisation error) for PINN-based RUL estimation under physical constraints.

8. Optimal common-cause failure prevention. The β-factor is treated as given in standard practice. But in system design, the investment in diversity and physical separation that reduces β is a decision variable. The tradeoff between CCF reduction cost and the associated reliability improvement (a non-linear function of β through the 1oo2 or 2oo3 PFD formula) defines an optimal β^*_{target}. *Open*: derive the optimal diversity investment strategy as a function of component costs, system architecture, and regulatory SIL targets.

A.1 Probability Distributions Used in This Volume

Distribution	Parameters	PDF / PMF	Mean / Variance
Exponential	$\lambda > 0$	$f(t) = \lambda e^{-\lambda t}$	$\mu = 1/\lambda$, $\sigma^2 = 1/\lambda^2$
Weibull	$k > 0, \lambda > 0$	$f(t) = k\lambda(\lambda t)^{k-1} e^{-(\lambda t)^k}$	$\mu = \Gamma(1 + 1/k)/\lambda$, $\sigma^2 = [\Gamma(1 + 2/k) - \Gamma(1 + 1/k)^2]/\lambda^2$
Log-Normal	$\mu \in \mathbb{R}, \sigma > 0$	$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} e^{-(\ln t - \mu)^2/(2\sigma^2)}$	$e^{\mu+\sigma^2/2}$, $(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$
Gamma	$\alpha, \beta > 0$	$f(t) = \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}$	$\mu = \alpha/\beta$, $\sigma^2 = \alpha/\beta^2$
Gompertz	$\lambda, \gamma > 0$	$f(t) = \lambda e^{\gamma t} \exp(-\frac{\lambda}{\gamma}(e^{\gamma t} - 1))$	No closed form
Log-Logistic	$\alpha, \beta > 0$	$f(t) = \frac{\alpha\beta(\alpha t)^{\beta-1}}{(1+(\alpha t)^\beta)^2}$	$\pi/(\alpha\beta \sin(\pi/\beta))$ for $\beta > 1$
Inverse Gaussian	$\mu, \lambda > 0$	$f(t) = \sqrt{\frac{\lambda}{2\pi t^3}} e^{-\lambda(t-\mu)^2/(2\mu^2 t)}$	μ, μ^3/λ
Beta	$a, b > 0$	$f(x) = x^{a-1}(1 - x)^{b-1}/B(a, b)$, $x \in [0, 1]$	$a/(a+b)$, $ab/[(a+b)^2(a+b+1)]$
Lomax	$\alpha, \lambda > 0$	$f(t) = \alpha\lambda(1 + \lambda t)^{-(\alpha+1)}$	$1/(\lambda(\alpha - 1))$ for $\alpha > 1$

A.2 Special Functions

Gamma function. $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$ for $\Re(z) > 0$. Recurrence: $\Gamma(z + 1) = z\Gamma(z)$. Key values: $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(n + 1) = n!$ for $n \in \mathbb{N}$.

Incomplete Gamma function. $\gamma(s, x) = \int_0^x t^{s-1} e^{-t}\, dt$ (lower), $\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t}\, dt$ (upper), with $\gamma(s, x) + \Gamma(s, x) = \Gamma(s)$. The regularised version $P(s, x) = \gamma(s, x)/\Gamma(s)$ is the CDF of the Gamma$(s, 1)$ distribution.

Digamma function. $\psi(z) = \Gamma'(z)/\Gamma(z) = -\gamma_E + \sum_{n=0}^\infty (1/(n + 1) - 1/(n + z))$ where $\gamma_E = 0.5772\ldots$ is Euler's constant. Recurrence: $\psi(z + 1) = \psi(z) + 1/z$. Appears in the gradient of the Gamma-process log-likelihood.

Error function and Mill's ratio. $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$, $\quad \Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}(x/\sqrt{2})\right]$. Mill's ratio: $m(x) = (1 - \Phi(x))/\phi(x) \approx 1/x$ for large x (used in log-normal hazard asymptotics).

Modified Bessel functions. $I_\nu(x) = \sum_{k=0}^\infty (x/2)^{2k+\nu}/(k!\,\Gamma(k+\nu+1))$. The Laplace transform of the Inverse Gaussian density involves $I_{-1/2}$; the MGF of the IG distribution involves K_ν (modified Bessel of second kind).

Laplace–Stieltjes transform. $\tilde{F}(s) = \int_0^\infty e^{-st}\, dF(t) = \mathbb{E}[e^{-sT}]$. The LST completely characterises F and appears in: renewal equation solutions ($M(s) = \tilde{F}(s)/(1 - \tilde{F}(s))$), mean ($\mathbb{E}[T] = -\tilde{F}'(0)$), and variance ($\operatorname{Var}[T] = \tilde{F}''(0) - [\tilde{F}'(0)]^2$).

A.3 Key Inequalities

Jensen's inequality. For convex g and random variable X: $g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)]$. Application: $\mathbb{E}[1/T] \geq 1/\mathbb{E}[T]$ (reciprocal is convex); average failure rate $\geq 1/\text{MTTF}$.

Markov's inequality. $\Pr(T \geq t) \leq \mathbb{E}[T]/t$. Crude but universal; tightened by Chebyshev ($\Pr(|T - \mu| \geq k\sigma) \leq 1/k^2$).

Mill's ratio bound. For $X \sim \mathcal{N}(0,1)$: $\phi(x)/x \cdot (1 - 1/x^2) \leq 1 - \Phi(x) \leq \phi(x)/x$. Critical for bounding log-normal tail probabilities and FORM approximations.

Chernoff bound. $\Pr(T \geq t) \leq e^{-st}\mathbb{E}[e^{sT}]$ for any $s > 0$. Optimising over s gives the sharpest exponential bound, underlying importance sampling and large deviations.

FKG inequality. For increasing events A, B on a lattice: $\Pr(A \cap B) \geq \Pr(A)\Pr(B)$. In system reliability: for monotone structure functions, $\Pr(\phi(\mathbf{X}) = 1) \geq \prod_i p_i$ (series lower bound is conservative for more-than-series systems).

A.4 Stochastic Processes and Martingales

Counting process formulation. Let $N(t) = \sum_{i=1}^{n} N_i(t)$ where $N_i(t) = \mathbf{1}[T_i \le t, \delta_i = 1]$ is the failure indicator for unit i. The compensated process $M_i(t) = N_i(t) - \int_0^t Y_i(s)h(s)\,ds$ is a martingale with respect to the natural filtration $\mathcal{F}_t$, where $Y_i(s) = \mathbf{1}[T_i \ge s]$ is the at-risk indicator. The *predictable variation process*:

$$\langle M_i \rangle(t) = \int_0^t Y_i(s)h(s)\,ds,$$

equals the cumulative intensity. The Nelson–Aalen estimator arises from the innovation theorem: $d\hat{H}(t) = dN(t)/Y(t)$ where $Y(t) = \sum_i Y_i(t)$.

Doob–Meyer decomposition. Every right-continuous submartingale $X(t)$ with $\mathbb{E}[\sup_t |X(t)|] < \infty$ decomposes uniquely as $X(t) = M(t) + A(t)$, where $M(t)$ is a martingale and $A(t)$ is a predictable increasing process. Applied to reliability: $N_i(t) = M_i(t) + \Lambda_i(t)$ with $\Lambda_i(t) = \int_0^t Y_i(s)h(s)\,ds$ (the compensator). This is the foundation of the Andersen–Gill model for recurrent events.

Brownian motion and its properties. Standard Brownian motion $\{W_t, t \ge 0\}$: (i) $W_0 = 0$ a.s.; (ii) independent increments; (iii) $W_t - W_s \sim \mathcal{N}(0, t-s)$ for $t > s$; (iv) continuous paths a.s. Key properties:

- *Quadratic variation*: $[W]_t = t$ a.s. (i.e. $\sum_k (W_{t_{k+1}} - W_{t_k})^2 \to t$ as mesh $\to 0$).
- *Reflection principle*: $\Pr(\max_{0 \le s \le t} W_s \ge b) = 2\Pr(W_t \ge b) = 2(1 - \Phi(b/\sqrt{t}))$ for $b > 0$.
- *Time inversion*: $\{tW_{1/t}\}$ is also a Brownian motion.
- *Scaling*: $\{c^{-1/2} W_{ct}\}$ is Brownian motion for any $c > 0$.

Itô's formula. For a smooth function $f(t, x)$ and $X(t) = \mu t + \sigma W_t$:

$$f(t, X(t)) = f(0, X(0)) + \int_0^t \frac{\partial f}{\partial s}\,ds + \int_0^t \frac{\partial f}{\partial x}\,dX(s) + \frac{\sigma^2}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}\,ds.$$

Applied to degradation: if $X(t) = \mu t + \sigma W_t$ and $f(x) = e^{-\theta x}$ (moment generating function), Itô's formula gives $\mathbb{E}[e^{-\theta X(t)}] = e^{-\theta\mu t + \theta^2\sigma^2 t/2}$ — confirming the Gaussian MGF.

Optional Stopping Theorem (OST). If $M(t)$ is a martingale and τ is a stopping time with $\mathbb{E}[\tau] < \infty$, then $\mathbb{E}[M(\tau)] = \mathbb{E}[M(0)]$. Applied to prove the IG moments: for $M(t) = W_t$ (martingale, $\mathbb{E}[M(0)] = 0$) and $\tau =$ FPT of $X(t) = \mu t + \sigma W_t$ to level ℓ: $W_\tau = (\ell - \mu\tau)/\sigma$, so $\mathbb{E}[(\ell - \mu\tau)/\sigma] = 0$, giving $\mathbb{E}[\tau] = \ell/\mu$. The second moment uses $M(t) = W_t^2 - t$.

A.5 Optimisation Theory for Maintenance Problems

Calculus of variations formulation. The continuous-time maintenance optimisation:

$$\min_{u(t) \in [0,1]} J[u] = \int_0^\infty e^{-rt}\big[c_{\text{op}} + c_{\text{maint}}\,u(t)\big]\,dt,$$

where $u(t)$ is the maintenance effort rate, subject to the degradation dynamics $\dot{X} = -u(t) \cdot r_{\text{restore}} + \mu_{\text{degrade}}$ and state constraint $X(t) \ge 0$. The Pontryagin maximum principle gives

necessary conditions: the optimal control is bang-bang ($u^*(t) \in \{0, 1\}$) when cost is linear in u, switching at the singular surface defined by the costate equation.

Convexity and the KKT conditions. The age replacement cost $g(T_p)$ is quasi-convex in T_p for IFR distributions: any local minimum is a global minimum. For the block replacement problem (discrete decision), the integer programming relaxation's KKT conditions recover the continuous optimum, which is then rounded. The Lagrangian relaxation of the RAP (Section 8.5) has a duality gap of zero when the reliability polynomial is concave in $\mathbf{n}$ (true for series-parallel systems).

Dynamic programming and the principle of optimality. Bellman's principle: the tail of an optimal policy is optimal for the reduced problem starting from any intermediate state. Formalised as: $V_t(\text{state}) = \min_a[c(a, \text{state}) + \mathbb{E}[V_{t+1}(\text{next state})|a]]$. Applied to maintenance: state = component age (or degradation level), action = do nothing, maintain, replace, horizon = mission length or infinite discounted. The optimal policy is characterised by the value function V^* satisfying the HJB equation (continuous time) or Bellman equation (discrete time).

Convex optimisation for model fitting. The Weibull, Cox, and Gamma-process log-likelihoods are all concave (negative Hessians positive definite), guaranteeing: (i) a unique global MLE; (ii) Newton–Raphson convergence from any starting point; (iii) well-defined Fisher information matrix $\mathcal{I} = -\mathbb{E}[\nabla^2 \ell]$ and asymptotic Cramér–Rao lower bound $\text{Var}[\hat{\theta}] \geq \mathcal{I}^{-1}$.

B.1 Proof: Elementary Renewal Theorem

Theorem. If $X_1, X_2, \ldots$ are i.i.d. with $0 < \mu = \mathbb{E}[X_i] < \infty$, then $M(t)/t \to 1/\mu$ as $t \to \infty$.

Proof. Since $S_{N(t)} \leq t < S_{N(t)+1}$, dividing by $N(t)$:

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)}.$$

By the SLLN: $S_n/n \to \mu$ a.s., so $S_{N(t)}/N(t) \to \mu$ a.s. (since $N(t) \to \infty$). Similarly $S_{N(t)+1}/N(t) = S_{N(t)+1}/(N(t)+1) \cdot (N(t)+1)/N(t) \to \mu \cdot 1 = \mu$ a.s. By the squeeze theorem: $t/N(t) \to \mu$ a.s., i.e. $N(t)/t \to 1/\mu$ a.s. Taking expectations and using dominated convergence (bounded convergence suffices for L^1): $M(t)/t \to 1/\mu$. ∎

B.2 Proof: Optimal Age Replacement Condition

Theorem. For an IFR distribution F with $C_p < C_f$, the optimal replacement age $T_p^* \in (0, \infty)$ exists and is unique, characterised by:

$$h(T_p^*) \int_0^{T_p^*} S(t)\, dt - F(T_p^*) = \frac{C_p}{C_f - C_p}.$$

Proof. Denote $\Psi(T) = h(T) \int_0^T S(t)\, dt - F(T)$. We show Ψ is strictly increasing, $\Psi(0) = 0$, and $\Psi(\infty) = \infty$ (for IFR distributions).

At zero: $\Psi(0) = h(0) \cdot 0 - 0 = 0$.

Derivative: $\Psi'(T) = h'(T) \int_0^T S(t)\, dt + h(T)S(T) - f(T) = h'(T) \int_0^T S(t)\, dt \geq 0$, with equality only if $h' = 0$ everywhere (exponential). For strictly IFR, $\Psi'(T) > 0$.

At infinity: For IFR, $h(t) \to \infty$ (non-DFR), so $h(T) \int_0^T S(t)\, dt \geq h(T) \int_0^1 S(t)\, dt \to \infty$, while $F(T) \leq 1$.

Therefore Ψ maps $(0,\infty)$ onto $(0,\infty)$ continuously and strictly increasingly, ensuring a unique solution T_p^* to $\Psi(T_p^*) = C_p/(C_f - C_p) > 0$.

To confirm T_p^* minimises $g(T_p)$: $g'(T_p) = 0 \Leftrightarrow \Psi(T_p) = C_p/(C_f - C_p)$, and $g''(T_p^*) > 0$ follows from $\Psi' > 0$. ∎

B.3 Proof: IFRA Closure Under Coherent Systems

Theorem (Barlow–Proschan, 1975). If $T_1, \dots, T_n$ are i.i.d. IFRA, then the lifetime $T_s = \phi(T_1, \dots, T_n)$ of any coherent system is IFRA.

Proof sketch. Step 1: For a series system of 2 i.i.d. IFRA units, $T_s = \min(T_1, T_2)$ with $S_s(t) = S(t)^2$. Since $-\ln S_s(t)/t = 2(-\ln S(t)/t)$ is non-decreasing (IFRA of T_i), T_s is IFRA. By induction, min of n i.i.d. IFRA is IFRA.

Step 2: For a parallel system, $T_s = \max(T_1, T_2)$ with $S_s(t) = 2S(t) - S(t)^2$. One shows $-\ln S_s(t)/t$ is non-decreasing by the log-concavity argument: $S_s(ct)/S_s(c't) \geq S_s(t)^{(c-c')/(1-c')}$ for $0 < c' < c < 1$ (Barlow–Marshall–Proschan, 1963).

Step 3: Any coherent system can be expressed as a series of parallel modules (minimal path set representation). IFRA is closed under series (Step 1) and parallel (Step 2) combinations, so the full system is IFRA by induction on the modular decomposition tree. ∎

B.4 Proof: Renewal Reward Theorem

Theorem. Let $\{(X_k, R_k)\}_{k \geq 1}$ be i.i.d. with $0 < \mu = \mathbb{E}[X_k] < \infty$ and $\mathbb{E}[|R_k|] < \infty$. Then:

$$\frac{\sum_{k=1}^{N(t)} R_k}{t} \xrightarrow{t \to \infty} \frac{\mathbb{E}[R]}{\mu} \quad \text{a.s.}$$

Proof. Write:

$$\frac{1}{t}\sum_{k=1}^{N(t)} R_k = \frac{N(t)}{t} \cdot \frac{1}{N(t)} \sum_{k=1}^{N(t)} R_k.$$

By the ERT: $N(t)/t \to 1/\mu$ a.s. By the SLLN: $\frac{1}{N(t)}\sum_{k=1}^{N(t)} R_k \to \mathbb{E}[R]$ a.s. (since $N(t) \to \infty$ a.s.). The product of two a.s. convergent sequences converges a.s.: $\frac{1}{t}\sum R_k \to \mathbb{E}[R]/\mu$ a.s. ∎

B.5 Proof: Memorylessness Uniqueness of the Exponential

Theorem. A non-negative continuous random variable T satisfies $\Pr(T > s + t | T > s) = \Pr(T > t)$ for all $s, t \geq 0$ if and only if $T \sim \text{Exp}(\lambda)$ for some $\lambda > 0$.

Proof. ($\Leftarrow$) Direct calculation: $\Pr(T > s + t | T > s) = e^{-\lambda(s+t)}/e^{-\lambda s} = e^{-\lambda t}$.

($\Rightarrow$) Let $G(t) = \Pr(T > t)$. The memoryless property states $G(s + t) = G(s)G(t)$ for all $s, t \geq 0$. This is Cauchy's functional equation on $[0, \infty)$ with G right-continuous and $G(0) = 1$. The only measurable solutions are $G(t) = e^{ct}$ for some constant c. Since $G(t) \leq 1$ for all $t \geq 0$ and $G(t) \to 0$ as $t \to \infty$ (otherwise $T = \infty$ a.s., excluded), we need $c < 0$. Setting $\lambda = -c > 0$ gives $G(t) = e^{-\lambda t}$. ∎

Statistical Tables

C.1 Weibull Quantiles

For $T \sim \text{Weibull}(k, \lambda)$, the p-th quantile is $t_p = (-\ln(1-p))^{1/k}/\lambda$.

$p \backslash k$	0.5	1.0	1.5	2.0	2.5	3.0	3.5
0.01	0.0001	0.0101	0.0469	0.1002	0.1513	0.1943	0.2298
0.05	0.0026	0.0513	0.1296	0.2265	0.3073	0.3688	0.4155
0.10	0.0111	0.1054	0.2144	0.3246	0.4194	0.4848	0.5314
0.20	0.0498	0.2231	0.3659	0.4729	0.5702	0.6360	0.6816
0.50	0.4805	0.6931	0.7830	0.8326	0.8630	0.8838	0.8987
0.80	2.590	1.6094	1.2505	1.1093	1.0394	0.9988	0.9734
0.90	5.302	2.3026	1.5832	1.3167	1.2060	1.1490	1.1138
0.95	9.491	2.9957	1.9028	1.5339	1.3871	1.3088	1.2612
0.99	21.49	4.6052	2.5594	1.9527	1.7399	1.6264	1.5553

Values shown are $t_p\lambda$ (dimensionless). Columns correspond to shape k; note $k = 1$ gives exponential quantiles $-\ln(1-p)$.

C.2 Critical Values for the Log-Rank Test

For testing $H_0 : S_1(t) = S_2(t)$ with m failures total, the log-rank statistic χ^2 follows χ^2_1 asymptotically.

Significance level α	$\chi^2_{1,\alpha}$	Equivalent $\lvert Z\rvert_\alpha$
0.10	2.706	1.645
0.05	3.841	1.960
0.025	5.024	2.241
0.01	6.635	2.576
0.001	10.828	3.291

For small samples ($m < 20$), use the exact permutation test.

C.3 Miner Damage Correction Factors

Recommended critical Miner damage values D_c (instead of 1.0) for different loading scenarios based on experimental evidence:

Loading sequence	Recommended D_c	Basis
High-to-low	0.5–0.7	Conservative (Corten–Dolan)
Low-to-high	1.2–2.0	Non-conservative
Random variable amplitude	0.7–1.0	Rainflow spectrum
Constant amplitude	1.0	Definition
Narrow-band Gaussian	1.0	NB approximation
Broad-band random	0.7–0.9	Dirlik correction

C.4 Standard Normal CDF (abridged)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

Key values: $\Phi(1.645) = 0.950$, $\Phi(1.960) = 0.975$, $\Phi(2.326) = 0.990$, $\Phi(2.576) = 0.995$, $\Phi(3.090) = 0.999$.

Notation Index

Symbol	Meaning	First defined
T	Random time to failure	Sec. 5.1
$F(t)$	CDF of T: $\Pr(T \le t)$	Sec. 5.1
$S(t)$	Survival function: $1 - F(t)$	Sec. 5.1
$f(t)$	Density: $F'(t)$	Sec. 5.1
$h(t)$	Hazard function: $f(t)/S(t)$	Sec. 5.1
$H(t)$	Cumulative hazard: $-\ln S(t)$	Sec. 5.1
$e(t)$	Mean residual life at age t	Sec. 5.1
μ	Mean failure time $\mathbb{E}[T]$	Sec. 5.1
δ_i	Censoring indicator (1=failure)	Sec. 5.4
n_j, d_j	Risk set size and failures at t_j	Sec. 5.4
$\hat{S}(t)$	Kaplan–Meier estimator	Sec. 5.4
$\hat{H}(t)$	Nelson–Aalen estimator	Sec. 5.4
$h_0(t)$	Baseline hazard (Cox model)	Sec. 5.5
$\boldsymbol{\beta}$	Regression coefficients (Cox)	Sec. 5.5
$\mathcal{R}(t)$	Risk set at time t	Sec. 5.5
S, N	Stress amplitude and cycles (fatigue)	Sec. 6.1
m, C	Basquin parameters	Sec. 6.1
S_e	Endurance (fatigue) limit	Sec. 6.1
D, D_c	Miner damage and critical value	Sec. 6.2
a, a_0, a_c	Crack length, initial, critical	Sec. 6.3
K, ΔK	Stress intensity factor and range	Sec. 6.3
K_{Ic}	Plane-strain fracture toughness	Sec. 6.3
C_P, m	Paris crack growth constants	Sec. 6.3
Y	Geometry factor (LEFM)	Sec. 6.3
$X(t)$	Degradation process	Sec. 6.4
ℓ	Failure threshold	Sec. 6.4
$\alpha(t), \beta$	Gamma process shape and rate	Sec. 6.4
$N(t)$	Renewal/failure count process	Sec. 7.1
$M(t)$	Renewal function $\mathbb{E}[N(t)]$	Sec. 7.1
μ_U, μ_V	Mean ON and OFF times	Sec. 7.1
T_p, T_p^*	Replacement age (and optimal)	Sec. 7.2
C_p, C_f	Planned and failure replacement cost	Sec. 7.2
$g(T_p)$	Long-run cost rate	Sec. 7.2
v_k	Virtual age after k-th repair	Sec. 7.4
q	Repair effectiveness (0=AGAN, 1=ABAO)	Sec. 7.4

Symbol	Meaning	First defined
$F_k(t)$	Cumulative incidence function, cause k	Sec. 7.5
$h_k(t)$	Cause-specific hazard	Sec. 7.5
$\phi(\mathbf{x})$	Structure function	Sec. 8.1
R_s	System reliability	Sec. 8.1
p_i	Component reliability	Sec. 8.1
$\mathcal{P}, \mathcal{K}$	Sets of minimal path/cut sets	Sec. 8.1
I_i^B	Birnbaum importance	Sec. 8.3
I_i^{FV}	Fussell–Vesely importance	Sec. 8.3
β	CCF beta-factor	Sec. 8.4
PFD	Probability of failure on demand	Sec. 8.4
$\mathbf{n}$	Redundancy allocation vector	Sec. 8.5
β_{HL}	Hasofer–Lind reliability index	App. C
$\Gamma(z)$	Gamma function	App. A
$\psi(z)$	Digamma function	App. A
$\phi(\cdot), \Phi(\cdot)$	Standard normal pdf and CDF	Throughout
$\tilde{F}(s)$	Laplace–Stieltjes transform of F	App. A

E.1 Copulas: Definition and Properties

Sklar's Theorem

For any joint CDF $F(t_1, \dots, t_n)$ with marginals $F_1(t_1), \dots, F_n(t_n)$, there exists a copula $C : [0,1]^n \to [0,1]$ such that:

$$F(t_1, \dots, t_n) = C(F_1(t_1), \dots, F_n(t_n)).$$

If all marginals are continuous, the copula C is unique. Conversely, for any copula C and marginals F_i, the above defines a valid joint CDF.

Fréchet–Hoeffding bounds. Every copula satisfies:

$$W(u_1, \dots, u_n) \le C(u_1, \dots, u_n) \le M(u_1, \dots, u_n),$$

where $M(u_1, \dots, u_n) = \min_i u_i$ (comonotonicity; perfect positive dependence) and $W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$ (for $n = 2$: countermonotonicity). For $n \ge 3$, W is not a valid copula. The independence copula $\Pi(u_1, \dots, u_n) = \prod_i u_i$.

Archimedean copulas. Defined by a generator $\varphi : [0,1] \to [0,\infty)$ (decreasing, convex, $\varphi(1) = 0$):

$$C(u_1, \dots, u_n) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_n)).$$

Key families:

Family	Generator $\varphi(t)$	$C(u,v)$	Kendall's τ
Clayton	$(t^{-\theta} - 1)/\theta$, $\theta > 0$	$\max(u^{-\theta} + v^{-\theta} - 1, 0)^{-1/\theta}$	$\theta/(\theta + 2)$
Gumbel	$(-\ln t)^{\theta}$, $\theta \ge 1$	$\exp(-((-\ln u)^{\theta} + (-\ln v)^{\theta})^{1/\theta})$	$1 - 1/\theta$
Frank	$-\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$	$-\frac{1}{\theta} \ln \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} + 1$	$1 - 4/\theta(1 - D_1(\theta))$

where $D_1(\theta)$ is the Debye function. Clayton exhibits lower tail dependence (failures cluster in the lower tail); Gumbel exhibits upper tail dependence; Frank is symmetric.

Tail dependence coefficients. The *upper tail dependence coefficient*:

$$\lambda_U = \lim_{u \to 1^-} \Pr(U_2 > u | U_1 > u) = \lim_{u \to 1^-} \frac{1 - 2u + C(u,u)}{1 - u}.$$

Lower tail: $\lambda_L = \lim_{u \to 0^+} C(u,u)/u$. For reliability:

- $\lambda_L > 0$: components tend to fail together at short times (CCF-like behavior).
- $\lambda_U > 0$: components tend to survive together to long times (common robust design).
- Gaussian copula: $\lambda_L = \lambda_U = 0$ (tail independence).
- Clayton: $\lambda_L = 2^{-1/\theta}$, $\lambda_U = 0$.
- Gumbel: $\lambda_L = 0$, $\lambda_U = 2 - 2^{1/\theta}$.

E.2 Copula-Based System Reliability

For a series system with n dependent components:

$$R_s(t) = \Pr(T_1 > t, \dots, T_n > t) = \bar{C}(S_1(t), \dots, S_n(t)),$$

where $\bar{C}(u_1, \ldots, u_n) = \Pr(U_1 > u_1, \ldots, U_n > u_n)$ is the survival copula. Under independence: $R_s = \prod_i S_i(t)$. Under comonotonicity (M-copula): $R_s = \min_i S_i(t)$ — the weakest component determines system life.

For a parallel system:

$$1 - R_s(t) = \Pr(T_1 \le t, \ldots, T_n \le t) = C(F_1(t), \ldots, F_n(t)).$$

Under independence: $1 - R_s = \prod_i F_i(t)$ (standard formula). Under countermonotonicity (W-copula for $n = 2$): $1 - R_s = \max(F_1 + F_2 - 1, 0)$ — the strongest component determines system life.

Bounds on system reliability. Under unknown dependence structure but known marginals:

$$\bar{C}^U(S_1, \ldots, S_n) \le R_s^{\text{series}} \le \min_i S_i(t),$$

where $\bar{C}^U$ is the Fréchet lower bound. The gap between bounds quantifies the *model risk* from dependence uncertainty.

Example — Copula-based CCF for two diesel generators

$T_1, T_2 \sim \text{Weibull}(k = 2, \mu = 5000\,\text{h})$, i.e. $S(t) = e^{-(t/\mu\Gamma(1.5))^2}$. At $t = 1000\,\text{h}$: $S(1000) = e^{-(1000/4431)^2} = e^{-0.0508} = 0.950$.
Series system (both must work): under independence: $R_s = 0.950^2 = 0.903$.
Under Clayton copula ($\theta = 2$, $\tau_K = 0.5$): $\bar{C}_{\text{Clayton},\theta=2}(0.950, 0.950)$. For survival Clayton: $\bar{C}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$ where $u = S_1, v = S_2$: $\bar{C} = (0.950^{-2} + 0.950^{-2} - 1)^{-1/2} = (1.108 + 1.108 - 1)^{-1/2} = (1.216)^{-1/2} = 0.907$.
Under Gumbel copula ($\theta = 2$, $\tau_K = 0.5$): $\bar{C}_{\text{Gumbel}}(0.950, 0.950) = \exp(-2(-\ln 0.950)^2)^{1/2} = \exp(-(0.0513)^2 \times 2)^{0.5}$. Wait — $-\ln(0.950) = 0.0513$; $\bar{C} = \exp(-(0.0513^\theta + 0.0513^\theta)^{1/\theta}) = \exp(-(2 \times 0.0513^2)^{1/2}) = \exp(-0.0725) = 0.930$.
Summary: $R_s^{\text{indep}} = 0.903 < R_s^{\text{Clayton}} = 0.907 \approx R_s^{\text{Gumbel}} = 0.930$. Positive dependence increases series system reliability here (both fail together, but if one survives both tend to survive).

E.3 Estimation of Copula Parameters

Inference Functions for Margins (IFM). Two-stage approach:

1. Fit marginals separately: $\hat{\theta}_i = \arg\max_{\theta_i} \sum_k \ln f_i(T_{ki}; \theta_i)$.
2. Transform to pseudo-uniform: $\hat{u}_{ki} = F_i(T_{ki}; \hat{\theta}_i)$.
3. Fit copula: $\hat{\alpha} = \arg\max_\alpha \sum_k \ln c(\hat{u}_{k1}, \ldots, \hat{u}_{kn}; \alpha)$, where $c = \partial^n C / \partial u_1 \cdots \partial u_n$ is the copula density.

IFM is consistent but less efficient than full MLE (joint optimisation over marginal and copula parameters simultaneously).

Rank-based estimation. The empirical copula:

$$\hat{C}(u_1, \ldots, u_n) = \frac{1}{m} \sum_{k=1}^{m} \prod_{i=1}^{n} \mathbf{1}[\hat{u}_{ki} \le u_i],$$

where $\hat{u}_{ki} = \hat{F}_i(T_{ki}) = \text{rank}(T_{ki})/m$ (empirical CDF). Kendall's τ sample estimator: $\hat{\tau} = \binom{m}{2}^{-1} \sum_{k<l} \text{sgn}(T_{k1} - T_{l1})\text{sgn}(T_{k2} - T_{l2})$. Map $\hat{\tau}$ to copula parameter via the closed-form $\tau(\alpha)$ relationship.

Goodness-of-fit for copulas. Cramér–von Mises statistic:

$$S_m = m \int_{[0,1]^n} [\widehat{C}(\mathbf{u}) - C(\mathbf{u}; \hat{\alpha})]^2 \, d\widehat{C}(\mathbf{u}),$$

approximated by $S_m = \sum_k [\widehat{C}(\hat{\mathbf{u}}_k) - C(\hat{\mathbf{u}}_k; \hat{\alpha})]^2$. p-values via parametric bootstrap: resample from $C(\cdot; \hat{\alpha})$, refit, recompute S_m.

E.4 Vine Copulas for High-Dimensional Dependence

For $n > 3$ components, bivariate copulas are combined in a graphical structure called a *vine*:

C-vine (canonical vine): One variable X_1 is central; all bivariate copulas $C_{1j}(F_1, F_j)$ are specified, then conditional copulas $C_{ij|1}(F_{i|1}, F_{j|1})$ layer the remaining dependence. An n-component C-vine has $\binom{n}{2}$ bivariate copula pairs.

D-vine (drawable vine): Components are arranged in a line; copulas are specified for neighbouring pairs, then conditional copulas for pairs separated by one component, and so on.

R-vine (regular vine): General tree structure; encompasses C- and D-vines as special cases. Fitted by the `VineCopula` package in R or `pyvinecopulib` in Python.

Application to multi-component system reliability. For a 5-component system with vine copula, the joint survival function $\Pr(T_1 > t, \ldots, T_5 > t)$ is evaluated by sequential conditional integration — intractable analytically but efficiently computed by Monte Carlo: sample 10^5 joint failure times from the vine, count the fraction exceeding t.

Example — D-vine copula for aircraft engine subsystems

Five aircraft engine subsystems: fan, compressor, combustor, turbine, exhaust. Pairwise dependencies estimated from $m = 340$ engine inspection records. D-vine pairs (Clayton, $\hat{\theta}_{ij}$):

- Fan–Compressor: $\hat{\theta}_{12} = 1.4$ ($\hat{\tau} = 0.41$) — strong.
- Compressor–Combustor: $\hat{\theta}_{23} = 0.8$ ($\hat{\tau} = 0.29$).
- Combustor–Turbine: $\hat{\theta}_{34} = 1.1$ ($\hat{\tau} = 0.35$).
- Turbine–Exhaust: $\hat{\theta}_{45} = 0.6$ ($\hat{\tau} = 0.23$).
- Conditional Fan–Combustor|Compressor: $\hat{\theta}_{13|2} = 0.3$ ($\hat{\tau} = 0.13$) — weak.

Series system 20,000-cycle reliability: $\hat{R}_s = 0.847$ (vine copula) vs. $R_s^{\text{indep}} = 0.871$. Positive dependence reduces series reliability: if one subsystem degrades, others are more likely to degrade too. The Clayton vine captures this lower-tail clustering correctly.

E.1 Martingales and Counting Processes

Counting Process Martingale Decomposition

Let $N(t)$ be a counting process (non-decreasing, right-continuous, integer-valued) with intensity $\lambda(t)$ (the conditional rate given the filtration $\mathcal{F}_{t^-}$). The *compensated process*:

$$M(t) = N(t) - \int_0^t \lambda(s)\, ds$$

is a local martingale. This is the Doob–Meyer decomposition: $N(t) = A(t) + M(t)$ (predictable compensator $A(t)$ plus martingale $M(t)$).

Why this matters for reliability.

- The Nelson–Aalen estimator $\hat{H}(t) = \sum_{t_j \le t} d_j/n_j$ is the nonparametric MLE of the compensator $A(t) = H(t)$ under right censoring. Its consistency follows from the martingale central limit theorem applied to $M(t)$.
- The log-rank test statistic is a stochastic integral $Z = \int_0^\infty W(t)\, dM(t)$ (where $W(t)$ is a weight process), asymptotically normal by the counting process CLT.
- The Cox partial likelihood is derived from the innovation of the counting process under the proportional hazards model.

Predictable variation. For a square-integrable martingale $M(t)$:

$$\langle M \rangle(t) = \int_0^t \lambda(s)\, ds \quad \text{(if } N(t) \text{ has unit jumps)},$$

and $M(t)^2 - \langle M \rangle(t)$ is a martingale. The predictable variation gives the asymptotic variance of counting process integrals:

$$\operatorname{Var}\left[\int_0^t W(s)\, dM(s)\right] \approx \int_0^t W(s)^2 \lambda(s)\, ds.$$

Rebolledo's CLT for counting processes. Under mild regularity conditions, if the predictable variation $\langle M_n \rangle(t) \to v(t)$ and the maximal jump $\sup_s |\Delta M_n(s)| \to 0$ in probability, then $M_n(t) \Rightarrow B(v(t))$ (Brownian motion at variance $v(t)$). This is the fundamental theorem enabling asymptotic normality of survival analysis estimators.

E.2 Markov Chains in Continuous Time

Continuous-Time Markov Chain (CTMC)

A CTMC $\{X(t), t \ge 0\}$ on state space $\mathcal{S}$ has transition rates $q_{ij} \ge 0$ $(i \ne j)$ with generator matrix Q: $Q_{ij} = q_{ij}$ for $i \ne j$, $Q_{ii} = -\sum_{j \ne i} q_{ij}$. The Kolmogorov forward equation:

$$\frac{d}{dt}\boldsymbol{\pi}(t) = \boldsymbol{\pi}(t)Q, \quad \boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)e^{Qt}.$$

Steady-state: $\boldsymbol{\pi}Q = \mathbf{0}$, $\boldsymbol{\pi}\mathbf{1} = 1$ (solve the linear system).

Matrix exponential computation. $e^{Qt} = \sum_{k=0}^{\infty} (Qt)^k/k!$. For small state spaces ($|\mathcal{S}| \leq 20$): eigendecomposition $Q = V\Lambda V^{-1}$, then $e^{Qt} = Ve^{\Lambda t}V^{-1}$. For large sparse Q: Krylov subspace methods (EXPOKIT library).

First-passage time distributions from CTMCs. For a CTMC with absorbing state ℓ (failure), partition states into transient $\mathcal{T}$ and absorbing $\{\ell\}$. Write $Q = \begin{pmatrix} Q_{TT} & \mathbf{q}_{T\ell} \\ \mathbf{0} & 0 \end{pmatrix}$. The time to absorption starting from state $i \in \mathcal{T}$ has:

$$\Pr(T_\ell > t \,|\, X(0) = i) = [\boldsymbol{\pi}_0 e^{Q_{TT}t}]_i,$$

and the matrix Q_{TT} (sub-generator) has all eigenvalues with negative real part. The failure time distribution is a *phase-type distribution* — a mixture of matrix-exponential terms.

Phase-type distributions. A random variable T is phase-type $\mathrm{PH}(\boldsymbol{\alpha}, Q_{TT})$ if it is the absorption time of a CTMC with initial distribution $\boldsymbol{\alpha}$ and sub-generator Q_{TT}. Key properties:

- $S(t) = \boldsymbol{\alpha} e^{Q_{TT}t}\mathbf{1}$.
- $f(t) = \boldsymbol{\alpha} e^{Q_{TT}t}(-Q_{TT}\mathbf{1})$.
- $\mathbb{E}[T] = \boldsymbol{\alpha}(-Q_{TT})^{-1}\mathbf{1}$.
- PH distributions are *dense* in the class of distributions on $[0, \infty)$: any distribution can be approximated arbitrarily closely by a PH distribution.

The Erlang, hyperexponential, and Coxian distributions are all special cases of PH distributions.

Example — Phase-type fit to a bimodal failure time distribution

Observed failure time distribution has two modes: early failures (mean 50h) and wear-out failures (mean 300h). Fit a 2-state Erlang + exponential mixture as $\mathrm{PH}(\boldsymbol{\alpha}, Q_{TT})$:

$$Q_{TT} = \begin{pmatrix} -(\mu_1 + q) & q \\ 0 & -\mu_2 \end{pmatrix}, \quad \boldsymbol{\alpha} = (\pi, 1-\pi),$$

with $\mu_1 = 1/50 = 0.02$, $\mu_2 = 1/300 = 0.00333$, $q = 0.1$ (transition from fast to slow mode), $\pi = 0.15$ (15% in fast-failure mode). Survival function:

$$S(t) = \pi e^{-\mu_1 t}\big[1 + \frac{q}{\mu_2 - \mu_1 - q}(1 - e^{-(\mu_2 - \mu_1 - q)t})\big] + (1-\pi)e^{-\mu_2 t}.$$

EM fitting of (π, μ_1, μ_2, q) from $n = 200$ failure times: $\hat{\pi} = 0.18$, $\hat{\mu}_1 = 0.019$, $\hat{\mu}_2 = 0.0031$, $\hat{q} = 0.12$. KS test: $D = 0.038$, $p = 0.91$ (excellent fit to bimodal data that neither Weibull nor log-normal can capture).

E.3 Lévy Processes and Jump Degradation

A *Lévy process* $\{X(t), t \geq 0\}$ has stationary independent increments and $X(0) = 0$ a.s. The Lévy–Khintchine representation:

$$\mathbb{E}[e^{i\theta X(t)}] = \exp\left(t\left(i\mu\theta - \frac{\sigma^2\theta^2}{2} + \int_{\mathbb{R}\setminus\{0\}} (e^{i\theta x} - 1 - i\theta x \mathbf{1}_{|x|\leq 1})\nu(dx)\right)\right),$$

where μ is drift, σ^2 is Gaussian component, and ν is the Lévy measure (governing jump sizes and rates).

For degradation modelling, we restrict to *subordinators* (non-decreasing Lévy processes, ν supported on $(0,\infty)$, no Gaussian component):

- *Gamma process*: $\nu(dx) = ce^{-\beta x}/x\,dx$ (infinite activity, continuous paths, all jumps infinitely small).
- *Compound Poisson*: $\nu(dx) = \lambda F_J(dx)$ (finite activity, discrete jumps of random size $J \sim F_J$, rate λ).
- *Inverse Gaussian process*: $\nu(dx) = (\lambda/(2\pi))^{1/2}x^{-3/2}e^{-\lambda(x-\mu)^2/(2\mu^2 x)/\mu}dx$ (arises as subordinator for Brownian motion with drift).
- *Tempered stable process*: $\nu(dx) = Ce^{-\alpha x}x^{-(1+\beta)}dx$ (generalises Gamma, has polynomial-decay jump tails).

First passage for subordinators. For a subordinator $X(t)$ with $\mathbb{E}[X(1)] = \mu_X$ and Laplace exponent $\Phi(s) = -\ln \mathbb{E}[e^{-sX(1)}]$, the FPT $T_\ell = \inf\{t : X(t) \geq \ell\}$ has LST:

$$\mathbb{E}[e^{-sT_\ell}] = e^{-\ell\Phi^{-1}(s)},$$

where Φ^{-1} is the right-continuous inverse of the Laplace exponent. For the Gamma process: $\Phi(s) = c\ln(1+s/\beta)$, so $\Phi^{-1}(s) = \beta(e^{s/c}-1)$.

E.4 Stochastic Differential Equations for Degradation

The general degradation SDE:

$$dX = \mu(X,t)\,dt + \sigma(X,t)\,dW_t + \int_{\mathbb{R}} z\,\tilde{N}(dt,dz),$$

where $\tilde{N}(dt,dz) = N(dt,dz) - \nu(dz)\,dt$ is the compensated jump measure. Three canonical forms:

Geometric Brownian motion (log-normal degradation): $dX = \mu X\,dt + \sigma X\,dW_t$, giving $X(t) = X(0)\exp((\mu - \sigma^2/2)t + \sigma W_t)$.

Cox–Ingersoll–Ross (mean-reverting, always positive): $dX = \kappa(\theta - X)\,dt + \sigma\sqrt{X}\,dW_t$. For $2\kappa\theta > \sigma^2$ (Feller condition), $X(t) > 0$ a.s. Stationary distribution: Gamma(α,β) with $\alpha = 2\kappa\theta/\sigma^2$, $\beta = 2\kappa/\sigma^2$.

Merton jump-diffusion: $dX = \mu\,dt + \sigma\,dW_t + J\,dN_t$ where $N_t \sim$ Poisson(λt) and $J \sim F_J$ (jump sizes). Models degradation with sudden large damage events (impact, overload) superimposed on continuous wear. FPT does not have a closed form in general; solved by Monte Carlo or Laplace inversion.

Example — CIR process for battery capacity fade

Battery capacity $X(t)$ (fraction of nominal) follows a CIR process: $dX = \kappa(\theta - X)\,dt + \sigma\sqrt{X}\,dW_t$ with $\kappa = 0.002$/cycle (mean reversion speed), $\theta = 0.85$ (long-run mean capacity), $\sigma = 0.005$ (diffusion). Initial capacity $X(0) = 1.0$. Failure threshold: $\ell = 0.80$.
Feller condition: $2\kappa\theta/\sigma^2 = 2 \times 0.002 \times 0.85/0.000025 = 136 \gg 1$ (satisfied easily).
Stationary distribution: Gamma$(136, 160)$ — very concentrated near 0.85.
The CIR model captures the observed mean-reverting behaviour of battery capacity: after deep discharge events, capacity partially recovers (mean reversion), but with stochastic fluctuations. The FPT $T_{0.80}$ is simulated via Euler–Maruyama ($\Delta t = 1$ cy-

cle): $X_{n+1} = X_n + \kappa(\theta - X_n)\Delta t + \sigma\sqrt{X_n}\sqrt{\Delta t}\, Z_n$ where $Z_n \sim \mathcal{N}(0,1)$.
Monte Carlo (10^4 paths): $\hat{\mathbb{E}}[T_{0.80}] = 1850$ cycles, $\hat{P}(T_{0.80} < 1000) = 0.023$.

Reliability Data Sources and Industry Standards

F.1 Component Failure Rate Databases

Reliability prediction requires baseline failure rate data. The principal sources are:

MIL-HDBK-217F (*Reliability Prediction of Electronic Equipment*, 1991): the US Department of Defense handbook providing failure rate models for electronic components as a function of temperature, voltage stress, environment (ground benign, airborne, naval, etc.), and quality level. Widely used despite known conservatism; superseded for new designs by IEC 61709 and FIDES.

IEC TR 62380 (formerly RDF 2000): French/European failure rate data for electronic components under commercial and industrial conditions. More optimistic than MIL-HDBK-217F; calibrated to field experience from Telecom, automotive, and industrial sectors.

FIDES Guide 2022: reliability prediction methodology developed by a consortium of French aerospace companies. Explicitly models the physics of failure for each component type and environmental loading (thermal, mechanical, humidity). Replaces MIL-HDBK-217F in European aerospace.

OREDA (*Offshore Reliability Data Handbook*, 6th ed., 2015): failure rate data for offshore oil and gas equipment (valves, compressors, pumps, pressure vessels, instruments) based on field experience from major North Sea operators. The most authoritative source for process industry reliability.

NPRD-2016 (*Nonelectronic Parts Reliability Data*): failure rates for mechanical, electromechanical, and structural components. Companion to MIL-HDBK-217F for non-electronic parts.

IEEE Std 493 (*IEEE Gold Book*, *Recommended Practice for the Design of Reliable Industrial and Commercial Power Systems*, 2007): failure and repair rate data for power system components (transformers, cables, circuit breakers, switchgear) based on a 1973–1990 IEEE survey of industrial plants.

T-Book (*Reliability Data of Components in Nordic Nuclear Power Plants*, 6th ed., 2005): failure rate data specific to BWR and PWR nuclear plant components, pooled from Swedish, Finnish, and other Nordic plants. Critical for nuclear PRA.

SPRINT (*Software Prediction and Reliability Tool*): reliability growth tracking for software systems, implementing the JM, MO, and Crow–AMSAA models.

F.2 Key Industry Standards

Standard	Title and Scope	Issuing Body
IEC 61508	Functional Safety of E/E/PE Safety-Related Systems	IEC
IEC 61511	Functional Safety for the Process Industry (SIS)	IEC
IEC 62061	Functional Safety of Machinery	IEC
ISO 13849	Safety of Machinery — Safety-Related Parts of Control Systems	ISO

Standard	Title and Scope	Issuing Body
EN 50129	Railway Applications — Safety-Related Electronic Systems	CENELEC
DO-178C	Software Considerations in Airborne Systems	RTCA
ARP4761	Guidelines for Safety Assessment Process (Aerospace)	SAE
MIL-STD-1629A	Failure Mode Effects and Criticality Analysis (FMECA)	US DoD
MIL-STD-882E	System Safety	US DoD
NUREG-1792	Good Practices for Implementing Human Reliability Analysis	US NRC
ISO 31000	Risk Management — Guidelines	ISO
EN 13306	Maintenance Terminology	CEN
ISO 55000	Asset Management — Overview, Principles and Terminology	ISO
ASTM E1820	Standard Test Method for Measurement of Fracture Toughness	ASTM
BS 7910	Guide to Methods for Assessing the Acceptability of Flaws in Metallic Structures	BSI
ASME Sec. XI	Rules for Inservice Inspection of Nuclear Power Plant Components	ASME
Eurocode 3 (EN 1993)	Design of Steel Structures (fatigue Part 1-9)	CEN
IIW Doc. XIII-2151	Recommendations for Fatigue Design of Welded Joints and Components	IIW

F.3 How to Use Reliability Data in Practice

Generic vs. specific data. Generic database failure rates (OREDA, MIL-HDBK-217F) are population averages that may not reflect the specific operating environment, maintenance quality, or design variant of the system being analysed. Plant-specific data (recorded failure events from CMMS/SCADA) should be used whenever statistically significant ($n_f \geq 5$ failures), combined with generic data via Bayesian updating (Section on Bayesian Reliability).

Environmental adjustment factors. Most databases provide base failure rates at a reference environment (25°C, benign). Multipliers for actual conditions:

- Temperature: Arrhenius factor $\exp[E_a/k_B(1/T_{\text{ref}} - 1/T_{\text{op}})]$.
- Vibration: power-law $(\text{g-level}/g_{\text{ref}})^{n_v}$ with $n_v \approx 2$–4.
- Humidity: Peck model $(H/H_{\text{ref}})^{n_H}$ with $n_H \approx 2.7$.
- Combined: $\lambda_{\text{adj}} = \lambda_{\text{base}} \times AF_T \times AF_V \times AF_H \times AF_{\text{quality}}$.

Uncertainty treatment. Database failure rates carry a *data uncertainty factor* (DUF or error factor EF): the 95th percentile is EF$\times$ the median. Typical EF:

- Generic databases (OREDA, MIL-HDBK-217F): EF $= 3$–10.
- Plant-specific data ($n_f \geq 10$): EF $= 2$–3.
- Component test data ($n_f \geq 50$): EF $= 1.3$–2.0.

In PRA, the failure rate uncertainty is propagated through the fault tree Monte Carlo to give a distribution on system failure probability (Section 8.2).

Extended Case Study: Level 1 PRA for a Simplified Nuclear Safety System

This appendix develops a complete quantitative PRA for a simplified Emergency Core Cooling System (ECCS), integrating fault tree analysis, importance measures, uncertainty propagation, and maintenance optimisation from Parts 5–8.

G.1 System Description

The ECCS protects the reactor during a Loss of Coolant Accident (LOCA) by injecting water into the core. The simplified system has:

- Two high-pressure injection (HPI) pumps (HPI-A, HPI-B) in parallel.
- One common injection header (IH) in series with the pump pair.
- Two check valves (CV-A, CV-B), each in series with its respective pump.
- One common power bus (PB) supplying both pump motors.
- An actuation logic system (ALS) that starts the pumps on demand.

Logical structure. Success requires:

$$\phi = \text{PB} \cdot \text{ALS} \cdot \text{IH} \cdot \big[(\text{HPI-A} \cdot \text{CV-A}) \cup (\text{HPI-B} \cdot \text{CV-B})\big].$$

G.2 Fault Tree Minimal Cut Sets

Top event: "ECCS fails to deliver flow on demand." Applying De Morgan: system fails if any of the following MCS occurs:

MCS	Components	Order
K_1	{PB}	1
K_2	{ALS}	1
K_3	{IH}	1
K_4	{HPI-A, HPI-B}	2
K_5	{HPI-A, CV-B}	2
K_6	{CV-A, HPI-B}	2
K_7	{CV-A, CV-B}	2

G.3 Quantitative Analysis

Basic event failure probabilities per demand (from T-Book and plant-specific data):

Component	ID	q_i	Error Factor
Power bus	PB	3.0×10^{-4}	5
Actuation logic	ALS	5.0×10^{-4}	3
Injection header	IH	2.0×10^{-4}	5
HPI pump (each)	HPI	1.5×10^{-3}	5
Check valve (each)	CV	8.0×10^{-4}	5

Minimal cut set probabilities:

$$
\begin{aligned}
&P(K_1) = 3.0 \times 10^{-4},\\
&P(K_2) = 5.0 \times 10^{-4},\\
&P(K_3) = 2.0 \times 10^{-4},\\
&P(K_4) = (1.5 \times 10^{-3})^2 = 2.25 \times 10^{-6},\\
&P(K_5) = P(K_6) = 1.5 \times 10^{-3} \times 8.0 \times 10^{-4} = 1.20 \times 10^{-6},\\
&P(K_7) = (8.0 \times 10^{-4})^2 = 6.4 \times 10^{-7}.
\end{aligned}
$$

System failure probability (rare event approximation):

$$P_f \approx \sum_{j=1}^{7} P(K_j) = 3.0{+}5.0{+}2.0{+}0.00225{+}0.0012{+}0.0012{+}0.00064 \approx 1.005{\times}10^{-3} \text{ per demand.}$$

The first-order cut sets (PB, ALS, IH) contribute $10^{-3}/10^{-3} \approx 99.4\%$ of the total. Second-order cut sets are negligible.

G.4 Importance Measures

Fussell–Vesely importances:

$$I^{FV}_{\text{ALS}} = P(K_2)/P_f = 5.0 \times 10^{-4}/1.005 \times 10^{-3} = 0.498.$$

ALS contributes nearly 50% of system unreliability alone. Similarly $I^{FV}_{\text{PB}} = 0.299$, $I^{FV}_{\text{IH}} = 0.199$. The HPI pumps and check valves each contribute $< 0.1\%$.

Risk Achievement Worth:

$$\text{RAW}_{\text{ALS}} = \frac{1 - R_s(0_{\text{ALS}})}{1 - R_s} = \frac{1}{P_f} = \frac{1.005 \times 10^{-3}}{5.0 \times 10^{-4} + \text{rest}} \approx \frac{1.005 \times 10^{-3}}{5.05 \times 10^{-4}} \approx 2.$$

Actually $\text{RAW}_{\text{ALS}} \approx P_f/(P_f - P(K_2) + 1 \times P_{\text{ALS failed}}) \approx 1.005 \times 10^{-3}/(5.0 \times 10^{-4} + 5.05 \times 10^{-4}) \approx 1.00{\times}10^{-3}/1.005{\times}10^{-3} \approx 1.0/0.497 \approx 2.0$ — if ALS fails, system failure probability doubles from 10^{-3} to $\approx 2 \times 10^{-3}$.

Risk Reduction Worth: $\text{RRW}_{\text{ALS}} = P_f/(P_f - P(K_2)) = 1.005 \times 10^{-3}/(5.05 \times 10^{-4}) = 1.99$ — making ALS perfect halves the total system failure probability.

Priority: Target ALS and PB improvements first (combined FV importance 80%).

G.5 CCF Analysis

Both HPI pumps are of the same design and are susceptible to CCF. With $\beta = 0.05$:

$$q^{\text{indep}}_{\text{HPI}} = (1 - \beta) \times 1.5 \times 10^{-3} = 0.95 \times 1.5 \times 10^{-3} = 1.425 \times 10^{-3},$$

$$q_{\text{CCF,HPI}} = \beta \times 1.5 \times 10^{-3} = 7.5 \times 10^{-5}.$$

New $P(K_4) = (1.425 \times 10^{-3})^2 + 7.5 \times 10^{-5} = 2.03 \times 10^{-6} + 7.5 \times 10^{-5} \approx 7.7 \times 10^{-5}$.

CCF of both pumps adds 7.5×10^{-5} to P_f, increasing the second-order contribution from 4.1×10^{-6} to 7.9×10^{-5} — a $19\times$ increase. However, the first-order cut sets still dominate ($P_f \approx 1.08 \times 10^{-3}$).

G.6 Uncertainty Propagation

All basic events are log-normal distributed with the given error factors (EF). Monte Carlo propagation (10^5 samples):

Results: P_f distribution is log-normal with:

- 5th percentile: 3.1×10^{-4}
- Median (50th): 1.0×10^{-3}
- 95th percentile: 3.3×10^{-3}

The 95th-percentile-to-median ratio (≈ 3.3) reflects the combined uncertainty of three single-component cut sets each with EF=3–5. The uncertainty range spans roughly one order of magnitude — typical for PRA with limited plant-specific data.

G.7 Optimal Maintenance Policy for ALS

The ALS is a dormant safety system: failures are only revealed at periodic proof tests (interval τ). With $\lambda_{\text{DU,ALS}} = 5 \times 10^{-4}$/demand $\approx 1/2000$ demands, and assuming exponential repair time $\mu_{\text{ALS}} = 24$/year (24 repairs/year if demanded):

$\text{PFD}_{\text{ALS}} \approx \lambda_{\text{DU,ALS}}\tau/2$. For $\tau = 6$ months (0.5 year): $\text{PFD}_{\text{ALS}} \approx (5 \times 10^{-4}/\text{demand} \times 2\ \text{demands/year}) \times 0.5/2 = 2.5 \times 10^{-4}$.

Proof test cost: $c_t = \$8{,}000$ per test. Each test induces a small failure probability $p_{\text{test}} = 0.001$ (test-induced failure). Long-run cost rate:

$$g(\tau) = c_t/\tau + c_f\lambda_{\text{DU,ALS}}\tau/2 + c_f p_{\text{test}}/\tau.$$

With $c_f = \$1{,}000{,}000$ (LOCA consequence):

$$g(\tau) = 8000/\tau + 500000\lambda_{\text{DU,ALS}}\tau + 1000000 \times 0.001/\tau.$$

$g(\tau) = (8000 + 1000)/\tau + 500000 \times 5 \times 10^{-4}\tau/\text{yr} = 9000/\tau + 250\tau$.

Minimising: $\tau^* = \sqrt{9000/250} = \sqrt{36} = 6$ months. Optimal — the current 6-month interval is confirmed as cost-optimal.

Summary of Volume II

This volume has developed the mathematical foundations of reliability theory and failure processes across four interconnected parts.

- **Part 5 (Survival Analysis and Hazard Functions):** We introduced the triad of survival function $S(t)$, hazard function $h(t)$, and cumulative hazard $H(t) = -\ln S(t)$. The three canonical parametric families — Exponential (memoryless, constant hazard), Weibull (power-law hazard, IFR/DFR by shape), and Log-Normal (unimodal hazard, multiplicative degradation) — were derived from first principles with full MLE and accelerated life testing extensions. The bathtub curve was formalised as a three-component Weibull mixture or a competing-risks sum of cause-specific hazards. Nonparametric estimation under censoring was covered via the Kaplan–Meier product-limit estimator and Nelson–Aalen cumulative hazard; the Cox proportional hazards model completed the regression framework via partial likelihood.

- **Part 6 (Fatigue, Fracture, and Degradation Mechanics):** Basquin's S-N law ($N = CS^{-m}$) and its statistical P-S-N generalisation were derived from energy dissipation arguments. Palmgren–Miner cumulative damage $D = \sum n_i/N_i$ was presented alongside its known non-conservative failure mode (high-to-low loading) and nonlinear corrections (Marco–Starkey, Corten–Dolan). Paris' law $da/dN = C_P(\Delta K)^m$ governing fatigue crack growth was integrated analytically to give closed-form remaining-life formulas. Process-based degradation models — Gamma process and Inverse Gaussian process — provided stochastic first-passage-time failure models with tractable MLE. The Wiener-process FPT formula (IG distribution) was derived via the reflection principle.

- **Part 7 (Renewal Theory and Maintenance Models):** The renewal equation $M(t) = F(t) + \int_0^t M(t-s)\,dF(s)$ and the elementary renewal theorem ($M(t)/t \to 1/\mu$) provided the asymptotic backbone. The renewal-reward theorem gave the long-run cost rate framework for age and block replacement policies; the optimal age T_p^* was characterised by the implicit hazard-rate equation, showing that PM is beneficial if and only if the hazard is IFR. Imperfect repair was formalised via Kijima's virtual age models (I and II), connecting repair quality q to the effective age of the component post-repair. Competing risks were treated with cause-specific versus subdistribution hazards, the Aalen–Johansen CIF estimator, and the fundamental identifiability result.

- **Part 8 (System Reliability and Redundancy):** Structure functions, minimal path/cut set duality, and the reliability polynomial were developed for series, parallel, k-out-of-n, and bridge systems. Fault tree analysis (AND/OR gates, minimal cut sets, rare event approximation) and reliability block diagrams provided both qualitative and quantitative system models. Birnbaum, criticality, and Fussell–Vesely importance measures ranked components by sensitivity, critcality, and CDF contribution; RAW and RRW supplied the nuclear-safety perspective. Common-cause failures — the dominant failure mechanism in redundant systems — were modelled via β-factor, α-factor, MGL, and copula approaches. The redundancy allocation problem (RAP) was solved by Lagrangian relaxation and dynamic programming, completing the quantitative design framework.

Part	Core Model	Key Result	Estimator	Design Tool
5	Hazard $h(t)$	$S(t) = e^{-H(t)}$	Kaplan–Meier	Cox PH regression
6	Paris crack growth	$N_f = \int_{a_0}^{a_c} da/C_P\Delta K^m$	Gamma/IG process MLE	Basquin + Miner
7	Renewal equation	$M(t)/t \to 1/\mu$	Virtual age MLE	Age replacement T_p^*
8	Structure function	$R_s = \prod_j R_j$ (series)	Importance measures	RAP (DP/MIP)

Bibliography

References

[1] Aalen, O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, **5**(3), 141–150.

[2] Anderson, T. L. (2017). *Fracture Mechanics: Fundamentals and Applications* (4th ed.). CRC Press.

[3] Barlow, R. E. and Proschan, F. (1996). *Mathematical Theory of Reliability*. SIAM Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia.

[4] Basquin, O. H. (1910). The exponential law of endurance tests. *Proceedings of the American Society for Testing Materials*, **10**, 625–630.

[5] Black, J. R. (1969). Electromigration failure modes in aluminum metallization for semiconductor devices. *Proceedings of the IEEE*, **57**(9), 1587–1594.

[6] Chaboche, J. L. (1988). Continuum damage mechanics: Part I — General concepts. *Journal of Applied Mechanics*, **55**(1), 59–64.

[7] Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**(2), 187–220.

[8] Crow, L. H. (1974). Reliability analysis for complex, repairable systems. In *Reliability and Biometry* (F. Proschan and R. J. Serfling, eds.), SIAM, Philadelphia, 379–410.

[9] Dirlik, T. (1985). *Application of Computers in Fatigue Analysis*. PhD thesis, University of Warwick.

[10] Dugan, J. B., Bavuso, S. J. and Boyd, M. A. (1992). Dynamic fault-tree models for fault-tolerant computer systems. *IEEE Transactions on Reliability*, **41**(3), 363–377.

[11] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II (2nd ed.). Wiley, New York.

[12] Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**(446), 496–509.

[13] Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

[14] Fleming, K. N. and Mosleh, A. (1985). *Classification and Analysis of Reactor Operating Experience Involving Dependent Events*. EPRI NP-3967, Electric Power Research Institute.

[15] Greenwood, M. (1926). The natural duration of cancer. *Reports on Public Health and Medical Subjects*, **33**, 1–26.

[16] Hjort, N. L. (1990). Nonparametric Bayes estimators based on Beta processes in models for life history data. *The Annals of Statistics*, **18**(3), 1259–1294.

[17] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.

[18] International Electrotechnical Commission (2010). *Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems*, IEC 61508, Parts 1–7 (2nd ed.). IEC, Geneva.

[19] Ishwaran, H. and Kogalur, U. B. (2008). Random survival forests for R. *R News*, **7**(2), 25–31.

[20] Jardine, A. K. S. and Tsang, A. H. C. (2013). *Maintenance, Replacement and Reliability: Theory and Applications* (2nd ed.). CRC Press.

[21] Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data* (2nd ed.). Wiley, Hoboken.

[22] Kijima, M. (1989). Some results for repairable systems with general repair. *Journal of Applied Probability*, **26**(1), 89–102.

[23] Lam, Y. (1988). A note on the optimal replacement problem. *Advances in Applied Probability*, **20**(2), 479–482.

[24] Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data* (2nd ed.). Wiley, Hoboken.

[25] Lemaitre, J. and Chaboche, J. L. (1990). *Mechanics of Solid Materials*. Cambridge University Press.

[26] Lisnianski, A. and Levitin, G. (2003). *Multi-State System Reliability: Assessment, Optimisation and Applications*. World Scientific, Singapore.

[27] Miner, M. A. (1945). Cumulative damage in fatigue. *Journal of Applied Mechanics*, **12**(3), A159–A164.

[28] Mosleh, A., Fleming, K. N., Parry, G. W., Paula, H. M., Worledge, D. H. and Rasmuson, D. (1998). *Guidelines on Modelling Common-Cause Failures in Probabilistic Risk Assessment*. NUREG/CR-5485, US Nuclear Regulatory Commission.

[29] Moubray, J. (1997). *Reliability-Centred Maintenance* (2nd ed.). Butterworth-Heinemann, Oxford.

[30] Nelson, W. (1980). Accelerated life testing – step-stress models and data analysis. *IEEE Transactions on Reliability*, **29**(2), 103–108.

[31] van Noortwijk, J. M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, **94**(1), 2–21.

[32] Paris, P. and Erdogan, F. (1963). A critical analysis of crack propagation laws. *Journal of Basic Engineering*, **85**(4), 528–534.

[33] Provan, J. W. (ed., 1987). *Probabilistic Fracture Mechanics and Reliability*. Martinus Nijhoff, Dordrecht.

[34] Ramirez-Marquez, J. E. and Coit, D. W. (2004). A heuristic for solving the redundancy allocation problem for multi-state series-parallel systems. *Reliability Engineering & System Safety*, **83**(3), 341–349.

[35] Rauzy, A. (1993). New algorithms for fault trees analysis. *Reliability Engineering & System Safety*, **40**(3), 203–211.

[36] Rice, J. R. (1968). A path independent integral and the approximate analysis of strain concentration by notches and cracks. *Journal of Applied Mechanics*, **35**(2), 379–386.

[37] Suresh, S. (1998). *Fatigue of Materials* (2nd ed.). Cambridge University Press.

[38] Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, **72**(1), 20–22.

[39] Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, **14**(3), 809–834.

[40] Vesely, W. E., Goldberg, F. F., Roberts, N. H. and Haasl, D. F. (1981). *Fault Tree Handbook*. NUREG-0492, US Nuclear Regulatory Commission, Washington DC.

[41] Whitmore, G. A. and Schenkelberg, F. (1997). Modelling accelerated degradation data using Wiener diffusion with a time scale transformation. *Lifetime Data Analysis*, **3**(1), 27–45.

[42] Wöhler, A. (1870). Über die Festigkeits-Versuche mit Eisen und Stahl. *Zeitschrift für Bauwesen*, **20**, 73–106.